

Cambridge Introductions to Language and Linguistics

word
mental lexicon
inflection

lexeme

Introducing **Morphology**

Rochelle Lieber

THIRD EDITION

dictionary
compound

morpheme

Introducing Morphology

THIRD EDITION

A lively introduction to morphology, this textbook is intended for undergraduates with relatively little background in linguistics. It shows students how to find and analyze morphological data and presents them with basic concepts and terminology concerning the mental lexicon, inflection, derivation, morphological typology, productivity, and the interfaces between morphology and syntax on the one hand and phonology on the other. By the end of the text students are ready to understand morphological theory and how to support or refute theoretical proposals. Providing data from a wide variety of languages, the text includes hands-on activities designed to encourage students to gather and analyze their own data. The third edition has been thoroughly updated with new examples and exercises. Chapter 2 now includes an updated detailed introduction to using linguistic corpora, and there is a new final chapter covering several current theoretical frameworks.

ROCHELLE LIEBER is Professor of Linguistics at the University of New Hampshire, Durham, NH, where she teaches a wide range of courses on theoretical linguistics and the English language. She is the recipient of a Teaching Excellence Award (1990), the Lindberg Award for Outstanding Teacher and Scholar in Liberal Arts at UNH (2013), and the Bloomfield Award given by the Linguistic Society of America for the *Oxford Reference Guide to English Morphology* (with Laurie Bauer and Ingo Plag, Oxford University Press, 2015). She is the author of four monographs and over fifty articles and book chapters on morphology and related topics, and is the co-editor of three handbooks on morphology.

Cambridge Introductions to Language and Linguistics

This textbook series provides students and their teachers with accessible introductions to the major subjects encountered within the study of language and linguistics. Assuming no prior knowledge of the subject, each book is written and designed for ease of use in the classroom or seminar, and is ideal for adoption on a modular course as the core recommended textbook. Each book offers the ideal introductory material for each subject, presenting students with an overview of the main topics encountered in their course, and features a glossary of useful terms, chapter previews and summaries, suggestions for further reading, and helpful exercises. Each book is accompanied by a supporting website.

Books published in the series

Introducing Speech and Language Processing John Coleman

Introducing Phonetic Science Michael Ashby and John Maidment

Introducing Second Language Acquisition, Second Edition, Muriel Saville-Troike

Introducing English Linguistics Charles F. Meyer

Introducing Semantics Nick Riemer

Introducing Language Typology Edith A. Moravcsik

Introducing Psycholinguistics Paul Warren

Introducing Phonology, Second Edition, David Odden

Introducing Morphology, Third Edition, Rochelle Lieber

Introducing Morphology

THIRD EDITION

ROCHELLE LIEBER
University of New Hampshire



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi –
110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108832489

DOI: [10.1017/9781108957960](https://doi.org/10.1017/9781108957960)

© Rochelle Lieber 2010, 2016, 2022

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2010

Second Edition 2016

9th printing 2019

Third Edition 2022

Printed in the United Kingdom by TJ Books Limited, Padstow, Cornwall

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication Data

Names: Lieber, Rochelle, 1954– author.

Title: *Introducing morphology* / Rochelle Lieber.

Description: Third edition. | Cambridge, UK ; New York : Cambridge University
Press, 2021. | Series: Cambridge introductions to language and linguistics | Includes
bibliographical references and index.

Identifiers: LCCN 2021013065 (print) | LCCN 2021013066 (ebook) | ISBN 9781108832489
(hardback) | ISBN 9781108957960 (ebook)

Subjects: LCSH: Grammar, Comparative and general – Morphology. | BISAC: LANGUAGE
ARTS & DISCIPLINES / Linguistics / Morphology

Classification: LCC P241 .L534 2021 (print) | LCC P241 (ebook) | DDC 415–dc23

LC record available at <https://lcn.loc.gov/2021013065>

LC ebook record available at <https://lcn.loc.gov/2021013066>

ISBN 978-1-108-83248-9 Hardback

ISBN 978-1-108-95848-6 Paperback

Additional resources for this publication at www.cambridge.org/lieber3

Cambridge University Press has no responsibility for the persistence or accuracy of
URLs for external or third-party internet websites referred to in this publication,
and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

Contents

<i>Preface to First Edition</i>	ix
<i>Preface to Second Edition</i>	xii
<i>Preface to Third Edition</i>	xiii
<i>The International Phonetic Alphabet</i>	xiv
<i>Point and Manner of Articulation of English Consonants and Vowels</i>	xv
1 What Is Morphology?	1
1.1 Introduction	2
1.2 What's a Word?	3
1.3 Words and Lexemes, Types and Tokens	4
1.4 But Is It <i>Really</i> a Word?	5
1.5 Why Do Languages Have Morphology?	5
1.6 The Organization of This Book	7
<i>Summary</i>	8
<i>Exercises</i>	9
2 Words, Dictionaries, and the Mental Lexicon	11
2.1 Introduction	12
2.2 Why Not Check the Dictionary?	13
2.3 The Mental Lexicon	15
2.4 More about Dictionaries	25
<i>Summary</i>	33
<i>Exercises</i>	34
3 Lexeme Formation: The Familiar	35
3.1 Introduction	36
3.2 Kinds of Morphemes	36
3.3 Affixation	39
3.4 Compounding	50
3.5 Conversion	58
3.6 Marvelous Intricacies: How Affixation, Compounding, and Conversion Interact	60
3.7 Minor Processes	60
3.8 How To: Finding Data for Yourself	63
<i>Summary</i>	66
<i>Exercises</i>	67
4 Productivity and Creativity	71
4.1 Introduction	72
4.2 Factors Contributing to Productivity	73
4.3 Restrictions on Productivity	76
4.4 Ways of Measuring Productivity	77

4.5	Historical Changes in Productivity	79
4.6	Productivity <i>versus</i> Creativity	82
	<i>Summary</i>	84
	<i>Exercises</i>	84
5	Lexeme Formation: Further Afield	87
5.1	Introduction	88
5.2	How To: Morphological Analysis	89
5.3	Affixes: Beyond Prefixes and Suffixes	91
5.4	Internal Stem Change	95
5.5	Reduplication	96
5.6	Templatic Morphology	98
5.7	Subtractive Processes	100
	<i>Summary</i>	101
	<i>Exercises</i>	101
6	Inflection	105
6.1	Introduction	106
6.2	Types of Inflection	106
6.3	Inflection in English	119
6.4	Paradigms	123
6.5	Inflection and Productivity	126
6.6	Inherent <i>versus</i> Contextual Inflection	127
6.7	Inflection <i>versus</i> Derivation Revisited	128
6.8	How To: More Morphological Analysis	130
	<i>Summary</i>	132
	<i>Exercises</i>	133
7	Typology	137
7.1	Introduction	138
7.2	Universals and Particulars: A Bit of Linguistic History	138
7.3	The Genius of Languages: What's in Your Toolkit?	139
7.4	Ways of Characterizing Languages	151
7.5	Genetic and Areal Tendencies	158
7.6	Typological Change	160
	<i>Summary</i>	160
	<i>Exercises</i>	161
8	Words and Sentences: The Interface between Morphology and Syntax	165
8.1	Introduction	166
8.2	Argument Structure and Morphology	166
8.3	On the Borders	172
8.4	Morphological <i>versus</i> Syntactic Expression	176
	<i>Summary</i>	178
	<i>Exercises</i>	178

9	Sounds and Shapes: The Interface between Morphology and Phonology	181
9.1	Introduction	182
9.2	Allomorphy and Morphophonological Rules	182
9.3	Other Morphology–Phonology Interactions	190
9.4	How To: Morphophonological Analysis	193
9.5	Lexical Strata	197
	<i>Summary</i>	201
	<i>Exercises</i>	201
10	Theoretical Challenges	207
10.1	Introduction	208
10.2	The Nature of Morphological Rules	210
10.3	Lexical Integrity	214
10.4	Blocking, Competition, and Affix Rivalry	217
10.5	Constraints on Affix Ordering	219
10.6	Bracketing Paradoxes	221
10.7	The Nature of Affixal Polysemy	224
10.8	Reprise: What’s Theory?	226
	<i>Summary</i>	226
	<i>Exercises</i>	226
11	Theories of Morphology	229
11.1	Introduction	230
11.2	Distributed Morphology	231
11.3	Construction Morphology	234
11.4	Paradigm Function Morphology	238
11.5	Natural Morphology	240
11.6	Naïve Discriminative Learning	243
11.7	The Lexical Semantic Framework	245
11.8	A Brief Final Meditation on Theories	248
	Further Reading	249
	<i>Glossary</i>	251
	<i>References</i>	265
	<i>Index</i>	273

Preface to First Edition

One of the things that drew me to linguistics several decades ago was a sense of wonder at both the superficial diversity and the underlying commonality of languages. My wonder arose in the process of working through my first few problem sets in linguistics, not surprisingly, problem sets that involved morphological analysis. What I learned first was not theory – indeed at that moment in linguistic history morphology was not perceived as a separate theoretical area in the US – but what languages were like, how to analyze data, and what to call things. I love morphological theory, but for drawing beginning students into the field of linguistics, I believe that there is no substitute for hands-on learning, and that is where this book starts.

This book is intended for undergraduate students who may have had no more than an introductory course in linguistics. It assumes that students know the International Phonetic Alphabet, and have a general idea of what linguistic rules are, but it presupposes little else in the way of sophistication or technical knowledge. It obviously assumes that students are English-speakers, and therefore the first few chapters concentrate on English, and to some extent on languages that are likely to be familiar to linguistics students from language study in high school and university. As the book progresses, I introduce data from many languages that will be ‘exotic’ to students, so that by the end of the book, they will have some sense of linguistic diversity, at least with respect to types of morphology.

There are some aspects of the content of this text that might seem unusual to instructors. The first is the attention to dictionaries in [Chapter 2](#). Generally, texts on linguistic morphology do not mention dictionaries, but I find that beginning students of morphology retain a reverence for dictionaries that

sometimes gets in the way of thinking about the nature of the mental lexicon and how word formation works.

Instructors can skip all or part of this chapter, but my experience is that it sets students on a good footing from the start, and largely eliminates their squeamishness about considering whether *incent* or *bovineness* or *organizationalize* or the like are ‘real’ words, even if we can’t find them in the dictionary.

Another section that might seem odd is the part of [Chapter 7](#) devoted to snapshot descriptions of five different languages. These also might be skipped over, but they serve two important purposes. One purpose is simply to expose students to what the morphology of a language looks like overall; much of what they’re exposed to in the rest of the book (and in most other morphology texts that I know of) are bits and pieces of the morphology of languages – a reduplication rule here, an inflectional paradigm there – but never the big picture. More importantly, having looked at the ‘morphological toolkits’ of several languages, students will be better prepared to understand both the traditional categories used in morphological typology and more recent means of classification.

The final thing that might strike instructors as unusual is that I largely hold off on introducing morphological theory until the last chapter. Clearly, no text is theory-neutral, and this text is no exception. It fits squarely in the tradition of generative morphology in the sense that I present morphology as an attempt to characterize and model the mental lexicon. I presuppose that there is much that is universal in spite of apparent diversity. And I believe that the ultimate aim of teaching students about morphology (indeed about any area of linguistics) is to expose them to what is at stake in trying to characterize the nature

of the human language capacity. Nevertheless I start by presenting morphological rules in as neutral a way as possible, and hold off on raising theoretical disputes until students have enough experience to understand how morphological data might support or refute theoretical hypotheses. In a sense I believe that students will gain a better understanding of theory if they already have the ability to find data and analyze it themselves. Therefore the bulk of the morphological theory will be found in the last chapter, where I have tried to pick a few theoretical debates and show how one might argue for or against particular analyses. Having read this chapter, students will be able to go on and tackle some of the texts that are intended for advanced undergraduates or graduate students.

Since one of my main goals in this text is to teach students to do morphology, there are a number of pedagogical features that set this book apart from other morphology texts. First, each chapter has one or more “Challenge” boxes. These occur at points in the text where students might take a breather from reading or class lecture and try something out for themselves. Challenge exercises are ideal for small teams of students – either outside of class, or as an in-class activity – to work on together. Some involve discussion, some analysis, some doing some work online or at the library. But all of them involve hands-on learning. Instructors can use them or skip them or assign them as homework instead of, or in addition to, the exercises at the ends of chapters. I have tried most of them myself as in-class activities, and have found that they get students excited, stimulate discussion, and generally give students the feeling of really ‘doing morphology’ rather than just hearing about it.

A second pedagogical feature that sets this book apart are the “How to” sections in [chapters 3, 5, 6, and 9](#). These are meant to give students tips on finding or working with data. Some students don’t need such tips; they have the intuitive ability to look at data and figure out what to do with it. But I’ve found over

years of teaching that there are some students who don’t have this knack, and who benefit enormously from being walked through a problem or technique systematically. The “How to” sections do this.

Instructors and students will also find what they would expect to find in any good text. First, there are several aids to navigating the text – chapter outlines and lists of key terms at the beginnings of chapters and brief summaries at the end, as well as a glossary of the terms that are highlighted in the text. A copy of the International Phonetic Alphabet is included at the beginning for easy reference. And each chapter has a number of exercises that allow students to practice what they’ve been exposed to.

A general point about examples in this text. Where I have cited data from different books, grammars, dictionaries, and scholarly articles, I have chosen to keep the glosses provided in the original source even if this results in some inconsistency in the use of abbreviations. In other words, slightly different abbreviations may occur in different examples (for instance, N or Neut for ‘neuter’). Although students may be confused by this practice at first, it does give them a taste of the linguistic ‘real world’. Any student going on and doing further work in morphology is bound to find exactly this sort of variation in the use of abbreviations in sources. My goal in this text is to bring students to the point where they are not only ready to confront morphological theory but also have the skills to begin to think independently about it, and perhaps to contribute to it.

This text has benefitted from the help of many people. I am grateful to John McCarthy and Donca Steriade for suggesting examples, to Charlotte Brewer for supplying me with statistics about citations in the *OED*, to Marianne Mithun for suggesting Nishnaabemwin as a polysynthetic language to profile, and to several classes of students at UNH both for serving as guinea pigs on early drafts and for supplying me with wonderful examples from their Word Logs. Thanks go as well to the College of Liberal

Arts at the University of New Hampshire for the funds to hire a graduate student assistant at a critical moment, and to Chris Paris for supplying assistance. I am especially grateful to several anonymous reviewers who made

excellent suggestions on the penultimate draft of the text. Finally, thanks are due as well to Andrew Winnard at Cambridge University Press for inviting me to write this text and for his patience in waiting for it.

Preface to Second Edition

The study of morphology keeps on changing. There are basics that every student linguist must learn, but for all linguists – student and grown-up alike – there are always new challenges, new ideas, new ways of finding data. Textbooks that stay the same for too long therefore run the risk of falling behind the times. Hence, the need for a second edition. This edition is not radically different from the previous one, but I have made some significant additions. Most importantly, I have introduced the use of corpora as tools for gathering data. [Chapter 3](#) introduces students to gathering data from corpora such as the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC) and to formulating hypotheses on the basis of their own data. Exercises throughout the book now make reference to corpus data. I have also added some “How to” sections, as well as new

Challenge boxes within chapters. I have added material on the interaction of affixation, compounding, and conversion ([Chapter 3](#)), subtractive processes ([Chapter 5](#)), evidentiality ([Chapter 6](#)), typological change ([Chapter 7](#)), periphrastic versus morphological expression ([Chapter 8](#)), and syllable structure in morphology ([Chapter 9](#)). Exercises and additional examples have been added throughout.

I wish to thank several anonymous Cambridge University Press reviewers for comments both before and after I wrote this edition, as well as Andrew Winnard for his support throughout. I especially want to thank students in the Fall 2012 section of my morphology class for their great word log words and the students in the Fall 2014 section for serving as guinea pigs, finding typos, and generally letting me know what needed to be fixed. You guys are the best!

Preface to Third Edition

As I wrote in the preface to the second edition, morphology is a field that keeps on changing, and any good text must keep up with those changes. As I learn new things and use this text in my own teaching I find things that work well, but also things that fall flat. I find new subjects to add and ways to help my students with kinds of analysis that some find difficult.

This edition is not radically different from the previous one, but I have updated sections on the mental lexicon, added some “How to” sections, changed and clarified some exercises, and added sections on subjects I’ve learned about since writing the second edition (overabundance! mirativity!). I have also made some minor stylistic changes. Although I continue to preserve the language names and transcription systems that are used in my sources, I have added family (genetic) affiliations for languages outside of Indo-European, using the designations of Ethnologue (www.ethnologue.com). Languages for which no family affiliation is given can be assumed to be Indo-European.

The most substantial difference between the second and third editions is the addition of [Chapter 11](#) in which I give brief outlines of six theoretical frameworks that are or have been important in the first decades of the

twenty-first century. This edition still presupposes little prior linguistic knowledge, and is still eminently suitable for undergraduates just embarking on linguistic studies. But for courses in morphology that aim to prepare students (advanced undergraduates, beginning graduate students) for further study of the subject, [Chapter 11](#) can serve as a bridge between basic and advanced study. Instructors who are offering basic courses for undergraduates can choose to skip [Chapter 11](#) or assign it as recommended or supplementary reading.

I am grateful for the suggestions of several anonymous readers and for the ongoing support of Andrew Winnard and his staff at Cambridge University Press. As always, I am especially grateful to my own students. Students in my 2018 morphology class first gave me the idea of adding [Chapter 11](#) (thanks Tom and Jonah!). Students in my 2020 class served as guinea pigs for this edition, letting me know where things that seemed perfectly clear to me were not so clear at all. If the directions in the exercises have improved over the previous edition, the reader can thank Alex, Alexa, Brenda, Caleb, Denali, Eivet, Jessica, Jillian, and Sam. And of course, I thank all of these students for the words they contributed to the class word log, some of which have found their way into this edition.

The International Phonetic Alphabet

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ɸ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
Q Bilabial	b Bilabial	ʼ Ejectives
 Dental	d Dental/alveolar	pʼ Bilabial
! Postalveolar	f Palatal	tʼ Dental/alveolar
≠ Palatoalveolar	g Velar	kʼ Velar
 Alveolar lateral	ŋ Uvular	sʼ Alveolar fricative

OTHER SYMBOLS

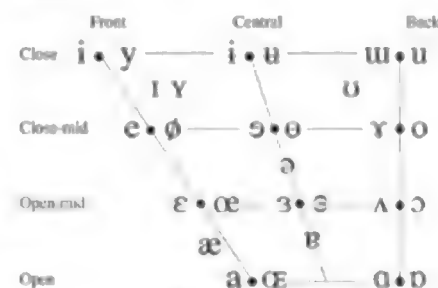
M	Voiced labiodental fricative	C Z	Alveolo-palatal fricatives
W	Voiced labiodental approximant	J	Voiced alveolar lateral flap
ɥ	Voiced labiodental approximant	ɟ	Simultaneous ʒ and x
H	Voiced epiglottal fricative		
ʕ	Voiced epiglottal fricative		
ʕ	Epiglottal pharynx		

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. $\dot{\eta}$

Voiceless	p	t	k	Breathy voiced	b	a	Dental	θ	ð
Voiceful	b	d	g	Creaky voiced	ɸ	ɶ	Apical	l	ɹ
Aspirated	p ^h	t ^h	k ^h	Linguallabial	ɸ	ɶ	Laminal	ɻ	ɽ
More rounded	ɔ	ʊ		Labiodental	t ^w	d ^w	Neutralized	ɛ	ə
Less rounded	ɔ	ʊ		Palatoalveolar	tʃ	dʃ	Nasal release	d ⁿ	
Advanced	u	ɪ		Velarized	t ^v	d ^v	Lateral release	d ^l	
Retracted	ɛ	ɔ		Pharyngealized	t ^ɕ	d ^ɕ	No audible release	d ^r	
Controlled	ɛ̃			Velarized or pharyngealized	t̠				
Mid-controlled	ẽ			Raised	e̥	(ɪ = voiced alveolar fricative)			
Syllabic	n			Lowered	e̜	(β = voiced bilabial approximant)			
Non-syllabic	ɱ			Advanced Tongue Root	e̟				
Phloetic	ɶ	a		Retracted Tongue Root	e̠				

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

	Primary stress
	Secondary stress
:	Long e:
ː	Half-long eː
˘	Extra-short ĕ
	Minor (foot) group
	Major (Intonation) group
.	Syllable break i.j.əkt
—	Linking (absence of a break)

TONES AND WORD ACCENTS
LEVEL CONTOL

LEVEL		CURVE	
e or ɛ	Extra high	e or ɛ	Rising
e	High	ẽ	Falling
ẽ	Mid	e	High-rising
e	Low	ẽ	Low-rising
ẽ	Extra low	e	Rising-falling
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

Point and Manner of Articulation of English Consonants and Vowels

Consonants								
	Labial	Labio-dental	Interdental	Alveolar	Alveo-palatal	Palatal	Velar	Glottal
Stop	p, b			t, d			k, g	ʔ
Fricative		f, v	θ, ð	s, z	ʃ, ʒ			h
Affricate					tʃ, dʒ			
Nasal	m			n			ŋ	
Liquid				l, r				
Glide	(w)					j	(w)	

Characters in boldface are voiced.

[w] is labio-velar in articulation.

Vowels			
	Front	Central	Back
High	i		u
	ɪ		ʊ
Mid	e	ʌ, ə	o
	ɛ		ɔ
Low	æ		ɑ

Tense vowels: **i, e, u, o, ɑ**

Lax vowels: **ɪ, ʊ, ɛ, æ, ʌ, ə**

Reduced vowel: **ə**

1 What Is Morphology?

CHAPTER OUTLINE

KEY TERMS

morpheme

word

simplex

complex

type

token

lexeme

lexeme (word)

formation

word form

inflection

In this chapter you will learn what morphology is, namely the study of word formation.

- ◆ We will look at the distinction between words and morphemes, between types, tokens, and lexemes, and between inflection and derivation.
- ◆ We will also consider the reasons why languages have morphology.

1.1 Introduction

The short answer to the question with which we begin this text is that morphology is the study of word formation, including the ways new words are coined in the languages of the world, and the way forms of words are varied depending on how they're used in sentences. As a native speaker of your language you have intuitive knowledge of how to form new words, and every day you recognize and understand new words that you've never heard before.

Stop and think a minute:

- Suppose that *splinch* is a verb that means 'step on broken glass'; what is its past tense?
- Speakers of English use the **suffixes** *-ize* (*crystallize*) and *-ify* (*codify*) to form verbs from nouns. If you had to form a verb that means 'do something the way Donald Trump does it', which suffix would you use? How about a verb meaning 'do something the way Joseph Biden does it'?
- It's possible to *rewash* or *reheat* something. Is it possible to *relove*, *reexplode*, or *rewiggle* something?

Chances are that you answered the first question with the past tense *splunched* (pronounced [splintʃt]),¹ the second with the verbs *Trumpify* and *Bidenize*, and that you're pretty sure that *relove*, *reexplode*, and *rewiggle* sound rather weird, if not downright unacceptable. Your ability to make up these new words, and to make judgments about words that you think could never exist, suggests that you have intuitive knowledge of the principles of word formation in your language, even if you can't articulate what they are. Native speakers of other languages have similar knowledge of their languages. This book is about that knowledge, and about how we as linguists can find out what it is. Throughout this book, you will be looking into how you form and understand new words, and how speakers of other languages do the same. Many of our examples will come from English – since you're reading this book, I assume we have that language in common – but we'll also look beyond English to how words are formed in languages with which you might be familiar, and languages which you might never have encountered before. You'll learn not only the nuts and bolts of word formation – how things are put together in various languages and what to call those nuts and bolts – but also what this knowledge says about how the human mind is organized.

The beauty of studying morphology is that even as a beginning student you can look around you and bring new facts to bear on our study. At this point, you should start keeping track of interesting cases of new words that you encounter in your life outside this class. Look at the first Challenge box.

1. In this text I presuppose that you have already learned at least that part of the International Phonetic Alphabet (IPA) that is commonly used for transcribing English. You'll find an IPA chart at the beginning of this book, if you need to refresh your memory.

Challenge: Your Word Log

Keep track of every word you hear or see (or produce yourself) that you think you've never heard before. You might encounter words while listening to the radio, watching TV, surfing the internet, or reading, or someone you're talking to might slip one in. Write those new words down, take note of where and when you heard/read/produced them, and jot down what you think they mean. What you write down may or may not be absolutely fresh new words – they just have to be new to you. We'll be coming back to these as the course progresses and putting them under the microscope.

Of course, if the answer to our initial question were as simple as the task in the box, you might expect this book to end right here. But there is of course much more to say about what makes up the study of morphology. Simple answers frequently lead to further questions, and here's one that we need to settle before we go on.

1.2 What's a Word?

Ask anyone what a word is and ... they'll look puzzled. In some sense, we all know what words are – we can list words of various sorts at the drop of a hat. But ask us to define explicitly what a word is, and the average non-linguist is flummoxed. One person might say that a word is a stretch of letters that occurs between blank spaces. But another is bound to point out that words don't have to be written for us to know that they're words. And in spoken (or signed) language, there are no spaces or pauses to delineate words. Yet we know what they are. Still another person might at this point try an answer like this: "A word is a small piece of language that means something," to which a devil's advocate might respond, "But what do you mean by 'a small piece of language'?" This is the point at which it becomes necessary to define a few specialized linguistic terms.

Linguists define a **morpheme** as the smallest unit of language that has its own meaning. Simple words like *giraffe*, *wiggle*, or *yellow* are morphemes, but so are prefixes like *re-* and *pre-* and suffixes like *-ize* and *-er*.² There's far more to be said about morphemes – as you'll see in later chapters of this book – but for now we can use the term morpheme to help us come up with a more precise and coherent definition of word. Let us now define a **word** as one or more morphemes that can stand alone in a language. Words that consist of only one morpheme, like the words in (1), can be termed **simple** or **simplex** words. Words that are made up of more than one morpheme, like the ones in (2), are called **complex words**:

2. In Chapter 2 we will give a more formal definition of prefix and suffix. For now it is enough to know that they are morphemes that cannot stand on their own, and that prefixes come before, and suffixes after, the root or main part of the word.

- (1) *Simplex words*
 - giraffe
 - fraud
 - murmur
 - oops
 - just
 - pistachio
- (2) *Complex words*
 - opposition
 - intellectual
 - crystallize
 - prewash
 - repressive
 - blackboard

We now have a first pass at a definition of what a word is, but as we'll see, we can be far more precise.

1.3 Words and Lexemes, Types and Tokens

How many words occur in the following sentence?

My friend and I walk to class together, because our classes are in the same building and we dislike walking alone.

You might have thought of at least two ways of answering this question, and maybe more. On the one hand, you might have counted every item individually, in which case your answer would have been 21. On the other hand, you might have thought about whether you should count the two instances of *and* in the sentence as a single word and not as separate words. You might even have thought about whether to count *walk* and *walking* or *class* and *classes* as different words: after all, if you were not a native speaker of English and you needed to look up what they meant in the dictionary, you'd just find one entry for each pair of words. So when you count words, you may count them in a number of ways.

Again, it's useful to have some special terms for how we count words. Let's say that if we are counting every instance in which a word occurs in a sentence, regardless of whether that word has occurred before or not, we are counting word **tokens**. If we count word tokens in the sentence above, we count 21. If, however, we are counting a word once, no matter how many times it occurs in a sentence, we are counting word **types**.

Counting this way, we count 20 types in the sentence above: the two tokens of the word *and* count as one type. A still different way of counting words would be to count what are called **lexemes**. Lexemes can be thought of as families of words that differ only in their grammatical endings or grammatical forms; singular, plural, and possessive forms of a noun (*class*, *classes*, *class's*, etc.), present, past, and participle forms of verbs (*walk*, *walks*, *walked*, *walking*), different forms of a pronoun (*I*, *me*, *my*, *mine*) each represent a single lexeme. One way of thinking about lexemes

is that they are the basis of dictionary entries; dictionaries typically have a single entry for each lexeme. So if we are counting lexemes in the sentence above, we would count *class* and *classes*, *walk* and *walking*, *I* and *my*, and *our* and *we* as single lexemes; the sentence then has 16 lexemes.

1.4 But Is It Really a Word?

In some sense we now know what words are – or at least what word types, word tokens, and lexemes are. But there's another way we can ask the question "What's a word?" Consider the sort of question you might ask when playing Scrabble: "Is *aalii* a word?" Or when you encounter an unfamiliar word: "Is *bouncebackability* a word?" What you're asking when you answer questions like these, is really the question "Is *xyz* a REAL word?" Our first impulse in answering those questions is to run for our favorite dictionary; if it's a real word it ought to be in the dictionary.

But think about this answer for just a bit, and you'll begin to wonder if it makes sense. Who determines what goes in the dictionary in the first place? What if dictionaries differ in whether they list a particular word? For example, the *Official Scrabble Player's Dictionary* lists *aalii* but not *bouncebackability*. The *Oxford English Dictionary Online* doesn't list *aalii*, but it does list *bouncebackability*. So which one is right? Further, what about words like *paralpinism*, *eruptionist*, or *schlumpadinka* that don't occur in any published dictionary yet, but can be encountered in the media? According to Word Spy (www.wordspy.com) *paralpinism* is a sport involving first climbing a mountain and then using a paraglider to get down; an *eruptionist* is a person who believes that the world will end in a huge volcanic explosion; and *schlumpadinka* is an adjective meaning 'unkempt'. And what about the word *schlumpadinkahood*, which I just made up? Once you know what *schlumpadinka* means, you have no trouble understanding my new word. If it consists of morphemes, has a meaning, and can stand alone, doesn't it qualify as a word according to our definition even if it doesn't appear in the dictionary?

What all these questions suggest is that we each have a **mental lexicon**, a sort of internalized dictionary that contains an enormous number of words that we can produce, or at least understand when we hear them. But we also have a set of word formation rules which allows us to create new words and understand new words when we encounter them. In the chapters to follow, we will explore the nature of our mental lexicon in detail, and think further about the "Is it really a word?" question. In answering this question we'll be led to a detailed exploration of the nature of our mental lexicon and our word formation rules.

1.5 Why Do Languages Have Morphology?

As native speakers of a language we use morphology for different reasons. We will go into both the functions of morphology and means of forming new words in great depth in the following chapters, but here, I'll just give you a taste of what's to come.

One reason for having morphology is to form new lexemes from old ones. We will refer to this as **lexeme formation**. (Many linguists use the term **word formation** in this specific sense, but this usage can be confusing, as all of morphology is sometimes referred to in a larger sense as ‘word formation’.) Lexeme formation can do one of three things. It can change the part of speech (or category) of a word, for example, turning verbs into nouns or adjectives, or nouns into adjectives, as you can see in the examples in (3):

- (3) *Category-changing lexeme formation*³
 V → N: amuse → amusement
 V → A: impress → impressive
 N → A: monster → monstrous

Some rules of lexeme formation do not change category, but they do add substantial new meaning:

- (4) *Meaning-changing lexeme formation*
 A → A: ‘negative A’ happy → unhappy
 N → N: ‘place where N lives’ orphan → orphanage
 V → V: ‘repeat action’ wash → rewash

And some rules of lexeme formation both change category and add substantial new meaning:

- (5) *Both category- and meaning-changing lexeme formation*
 V → A: ‘able to be Ved’ wash → washable
 N → V: ‘remove N from’ louse → delouse

Why have rules of lexeme formation? Imagine what it would be like to have to invent a wholly new word to express every single new concept. For example, if you wanted to talk about the process or result of amusing someone, you couldn’t use *amusement*, but would have to have a term like *zorch* instead. And if you wanted to talk about the process or result of resenting someone, you couldn’t use *resentment*, but would have to have something like *plitz* instead. And so on. As you can see, rules of lexeme formation allow for a measure of economy in our mental lexicons: we can recycle parts to come up with new words. It is probably safe to say that all languages have some ways of forming new lexemes, although, as we’ll see as this book progresses, those ways might be quite different from the means we use in English.

On the other hand, we sometimes use morphology even when we don’t need new lexemes. For example, we saw that each lexeme can have a number of **word forms**. The lexeme WALK has forms like *walk*, *walks*, *walked*, *walking* that can be used in different grammatical contexts. When we change the form of a word so that it fits in a particular grammatical context, we are concerned with what linguists call **inflection**. Inflectional word formation is word formation that expresses grammatical distinctions like number (singular vs. plural); tense (present vs.

3. The notation V → N means ‘changes a verb to a noun’.

past); person (first, second, or third); and case (subject, object, possessive), among others. It does not result in the creation of new lexemes, but merely changes the grammatical form of lexemes to fit into different grammatical contexts (we will look at this in detail in [Chapter 6](#)).

Interestingly, languages have wildly differing amounts of inflection. English has relatively little inflection. We create different forms of nouns according to number (*wombat*, *wombats*); we mark the possessive form of a noun with *-s* or *-s'* (*the wombat's eyes*). We have different forms of verbs for present and past and for present and past participles (*sing*, *sang*, *singing*, *sung*), and we use a suffix *-s* to mark the third person singular of a verb (*she sings*).

However, if you've studied Latin, Russian, ancient Greek, or even Old English, you'll know that these languages have quite a bit more inflectional morphology than English does. Even languages like French and Spanish have more inflectional forms of verbs than English does.

But some languages have much less inflection than English does. Mandarin Chinese (Sino-Tibetan), for example, has almost none. Rather than marking plurals by suffixes as English does, or by prefixes as the Niger-Congo language Swahili does, Chinese does not mark plurals or past tenses with morphology at all. This is not to say that a speaker of Mandarin cannot express whether it is one giraffe, two giraffes, or many giraffes that are under discussion, or whether the sighting was yesterday or today. It simply means that to do so, a speaker of Mandarin must use a separate word like *one*, *two* or *many* or a separate word for *past* to make the distinction.

(6) Wo jian guo yi zhi chang jing lu.
I see PAST one CLASSIFIER giraffe⁴

(7) Wo jian guo liang zhi chang jing lu.
I see PAST two CLASSIFIER giraffe

The word *chang jing lu* 'giraffe' has the same form regardless of how many long-necked beasts are of interest. And the verb 'to see' does not change its form for the past tense; instead, the separate word *guo* is added to express this concept. In other words, some concepts that are expressed via inflection in some languages are expressed by other means (word order, separate words) in other languages.

1.6 The Organization of This Book

In what follows, we'll return to all the questions we've raised here. In [Chapter 2](#), we'll revisit the question of what a word is, by further probing the differences between our mental lexicon and the dictionary, and look further into questions of what constitutes a "real" word. We'll look at the ways in which word formation goes on around us all the time,

4. We will explain in [Chapter 6](#) what we mean by classifier. For now it is enough to know that classifiers are words that must be used together with numbers in Mandarin.

and consider how children (and adults) acquire words, and how our mental lexicons are organized so that we can access the words we know and make up new ones. In [Chapter 3](#), we'll get down to the work of looking at some of the most common ways that new lexemes are formed: by adding prefixes and suffixes, by making up compound words, and by changing the category of words without changing the words themselves. In this chapter we'll concentrate on how words are structured in terms of both their forms and their meanings. Many of our examples will be taken from English, but we'll also look at how these kinds of word formation work in other languages. [Chapter 4](#) takes up a related topic, productivity: some processes of word formation allow us to form many new words freely, but others are more restricted. In this chapter we'll look at some of the determinants of productivity, and how productivity can be measured. [Chapter 5](#) will also be concerned with lexeme formation, but with kinds of lexeme formation that are less familiar to speakers of English. We'll look at forms of affixation that English does not have (infixation, circumfixation), processes like reduplication, and templatic morphology. Our focus will be on learning to analyze data that might on the surface seem to be quite unfamiliar. In [Chapter 6](#) we will turn to inflection, looking not only at the sorts of inflection we find in English, but also at inflectional systems based on different grammatical distinctions than we find in English, and systems that are far more complex and intricate. [Chapter 7](#) will be devoted to the subject of typology, different ways in which the morphological systems of the languages of the world can be classified and compared to one another. We'll look at some traditional systems of classification, as well as some that have been proposed more recently, and assess their pros and cons. [Chapters 8 and 9](#) will explore the relationship between the field of morphology and the fields of syntax on the one hand and phonology on the other. [Chapter 10](#) will introduce you to some of the interesting theoretical debates that have arisen in the field of morphology over the last two decades. Finally, [Chapter 11](#) introduces you to six current theoretical frameworks; learning about these will prepare you to do more advanced work in morphology.

Summary

Morphology is the study of words and word formation. In this chapter we have considered what a word is and looked at the distinction between word tokens, word types, and lexemes. We have divided word formation into derivation – the formation of new lexemes – and inflection, the different grammatical word forms that make up lexemes.

Exercises

1. Are the following words simple or complex?

- a. members
- b. prioritize
- c. handsome
- d. fizzy
- e. dizzy
- f. grammar
- g. writer
- h. rewind
- i. reject
- j. alligator

If you have difficulty deciding whether particular words are simple or complex, explain why you find them problematic.

2. Do the words in the following pairs belong to the same lexeme or to different lexemes?

- a. revolve revolution
- b. revolution revolutions
- c. revolve dissolve
- d. go went
- e. wash rewash

3. Count word tokens, types, and lexemes for this sentence and answer the questions below.

I've just replaced my old printer with a new one that prints much faster.

tokens _____

types _____

- a. Is *I've* one lexeme or two? Explain.
- b. Do *printer* and *print* belong to the same lexeme?

4. For the sentence below, first count word tokens and word types, and then answer the two questions below:

I will say now, just as I said yesterday, that the price of pickles is high but the price of pickled string beans is higher.

tokens _____

types _____

- a. Are *pickles* and *pickled* members of the same lexeme? Explain your answer.
- b. Is *string bean* one lexeme or two? Explain your answer.

5. What words belong to the same word family or lexeme as *sing*?

2 Words, Dictionaries, and the Mental Lexicon

CHAPTER OUTLINE

KEY TERMS

word

mental lexicon

lexicography

*the Gavagai
problem*

fast mapping

aphasia

reaction time

*storage versus
rules*

acquisition

In this chapter you will learn why we make a basic distinction between the dictionary and the mental lexicon.

- ◆ We will look at how linguists study the mental lexicon and how children acquire words.
- ◆ We will consider whether complex words are stored in the mental lexicon, or derived by rules, or both.
- ◆ And we will look further at how dictionaries have evolved and how they differ from one another and from the mental lexicon.

2.1 Introduction

In the [last chapter](#), we raised the question “What’s a word?” And we saw in [Section 1.2](#) that this question actually subsumes two more specific questions. In this chapter we will look more closely at those questions.

On the one hand, when we ask “What’s a word?” we may be asking about the fundamental nature of wordhood – as we saw, a far thornier philosophical question than it would seem at first blush. Native speakers of a language seem to know intuitively what a ‘word’ is in their language, even if they have trouble coming up with a definition of ‘word’. Interestingly, the *Oxford American Dictionary* seems to bank on this intuitive knowledge when it defines a word as “a single distinct meaningful element of speech or writing, used with others (or sometimes alone) to form a sentence and typically shown with a space on either side when written or printed.” We’ve already debunked part of the OAD definition: languages need not be written, but they still have words, and words don’t have blank space between them in spoken language. Nevertheless, the OAD’s definition works for most people: most dictionary users probably do not know the word *morpheme*, which we used in our definition of *word* in the [last chapter](#), but the OAD relies on the likelihood that they will not first think of something like the prefix *re-* as a single meaningful element, or something like *irniarualiunga* which means ‘I am making a doll’ in Central Alaskan Yup’ik (Eskimo-Aleut) ([Mithun 1999](#): 203), and constitutes not only a word, but also a whole sentence. In other words, the OAD’s definition works because dictionary users already have an intuitive idea of what a word is!

Morphologists, however, have the luxury of being more precise: we can define a word as a sequence of one or more morphemes that can stand alone in a language. But in doing so, we have not exhausted what’s interesting about our question.

Indeed, in [Chapter 1](#) we saw that there is a second way of interpreting it, one that seems far more concrete at first: we can interpret our question as meaning “Is xyz a word?” where xyz is a specific morpheme or sequence of morphemes. Taken this way, our question asks what it means to say that xyz is a word of English, or Central Alaskan Yup’ik, or some other language. On the one hand, we are always making up new words, and when we say them, others understand what we mean. In the [last chapter](#), I mentioned the words *eruptionist* and *schlumpadinka*, neither of which is in a (conventional) dictionary, at least as of the writing of this chapter, but both of which have been used (at least by me!). Does this qualify them as words? And two paragraphs up, I used the word *wordhood*, which you may or may not like, but which you certainly understood. This is the version of the “What’s a word?” question that we’ll concentrate on in this chapter. In doing so we’ll begin to explore the nature of dictionaries, and more importantly of our native speaker knowledge of words, which we might term our mental lexicon.

2.2 Why Not Check the Dictionary?

When the question “Is xyz really a word?” comes up – whether in casual conversation, in reading an article in the newspaper, or in playing Scrabble – people will often look to the dictionary for an answer. Which dictionary, of course, depends on what’s lying around the house or the office, or these days, what’s available online. But is this the right way to answer our question? As morphologists, we need to think about how dictionaries come to be, and how much we credit them with the authority to decide what’s a word.

There’s a lot to be said about how dictionaries have evolved and how they are produced today. For a short history of English dictionaries, you can read [Section 2.4](#) of this chapter. But for our immediate purposes, we can identify a number of reasons why we wouldn’t always want to base the answer to our question on what we find (or don’t) in a dictionary. Here are a few such reasons.

2.2.1 Which Dictionary?

Dictionaries come in all shapes and sizes, for all sorts of intended audiences. Size and audience are determined by individual publishers, and indeed the finished product is shaped by all sorts of market forces. And makers of dictionaries – **lexicographers** – are of course human; what gets into dictionaries has historically been subject to the individual foibles of lexicographers, not to mention the mores of society. If you grew up when I did, it was typical for dictionaries not to have taboo words like *fuck*, much less its derivatives *fuck*ing, *fuck* up, *fuck*able, *fuck* all, and *fuck*er, all of which can be found today in the *Concise Oxford English Dictionary*; but until the 1970s, dictionaries avoided words that might offend. It is perhaps safe to say that individual or societal foibles play less of a role in dictionary-making today, but it’s still a good idea to keep in mind that neither lexicographers nor the dictionaries they create are infallible.

Our first problem with giving final authority for wordhood to the dictionary, then, follows from the very concrete and temporal nature of dictionaries: if you look up a word in a pocket dictionary, or even a standard college desk dictionary, and it isn’t listed, you might still have the nagging suspicion that a bigger dictionary or a more specialized dictionary might list the word. But even if you check the largest available dictionary – say, for English the *Oxford English Dictionary Online* – or the most complete technical dictionary in a particular field, can you be sure that a word that’s not listed isn’t a word? Maybe it’s too new a word to have gotten into the dictionary yet.

2.2.2 Nonces, Mistakes, and Mountweazels

Further, sometimes we find items in dictionaries that we might hesitate to call words – even if they do occur in the dictionary. Among these items are words that are labeled as ‘**nonce**’, meaning that they’ve been found just once, often in the writing of someone

important, but that nevertheless don't seem to occur anywhere else. The *OED Online*, for example, lists as a nonce the word *agreemony*, which they define as 'agreeableness', and illustrate with a single quotation from the seventeenth-century writer Aphra Behn. Was this ever really a word? Indeed, the *OED* even lists some words that occur only once, and further in contexts which don't illuminate their meaning; for example, we can find the word *umbershoot* used by James Joyce in *Ulysses*, about which the *OED* maddeningly says only "meaning obscure"! Words or not?

Very extensive dictionaries like the *OED* sometimes also contain words that they identify as mistakes. For example, we can find an entry for the word *ambassady*, which occurs in a single quotation from 1693 and is, according to the *OED*, perhaps a mistake, where the author might have meant the word *ambassade* "the mission or function of an ambassador." It occurs in the dictionary, but is it really a word?

And finally, there are what have come to be called 'mountweazels'. A mountweazel is a phony word that is inserted into a dictionary so that its makers can identify lexicographic piracy. You can find a fuller explanation of this tradition in [Section 2.4](#), but the short version is this: lexicographers sometimes make up an entry and include it so that they can tell if another lexicographer is using their dictionary as a source without attribution (which is plagiarism, of course). Surely we wouldn't want to count such impostors as real words, but they're in the dictionary!

2.2.3 And the Problem of Complex Words

We will learn much more about this in the chapters to come, but perhaps the worst problem for us with the idea of giving the dictionary the authority to determine whether *xyz* is a word is that dictionaries don't need to include every word. Every language has ways of forming new words that are so active and transparent that putting all the words formed that way into the dictionary would be a waste of space. For example, speakers of English know that any verb at all can have a form made with the suffix *-ing*. As soon as I make up a new verb, say *zax*, we know that the *-ing* verb form is *zaxing*. So although a dictionary might eventually have to include the verb *zax*, it might never list *zaxing* as a word. But of course *zaxing* should be considered a word. Similarly, just about any adjective in English can be made into a noun by adding the suffix *-ness*. For example, the *Concise Oxford English Dictionary* contains the adjective *bovine*, but not the noun *bovineness*. Nevertheless, I'd have no problem if I saw the word *bovineness* written somewhere, and would never think to look it up in the dictionary. The dictionary doesn't have the word precisely because we'd never need to look it up.

The conclusion that we are inexorably led to is that we cannot rely on dictionaries to answer the question "Is *xyz* a word?" On the one hand, dictionaries don't list all the words of any language. They can't list all derivatives with living prefixes and suffixes, or all technical, scientific, regional, or slang words. And on the other hand, they sometimes include

words used only once whose meanings are completely unknown. They occasionally even include purposely made-up words to guard their own copyrights. For the most part, dictionaries do not fix or codify the words of a language, but rather reflect the words that native speakers use. Those words are encoded in what we will call the mental lexicon, the sum total of word knowledge that native speakers carry around in their heads. So to answer our question, we must look more closely at what is in that mental lexicon.

2.3 The Mental Lexicon

By the mental lexicon I mean the sum total of everything individual speakers know about the words of their language. This knowledge includes information about pronunciation, category (part of speech), and meaning, of course, but also information about syntactic properties (e.g., whether a verb is transitive or intransitive¹), level of formality, and what lexicographers call ‘range of application’, that is, the specific conditions under which we might use the word. For example, I know that the word *verandah* is a noun, pronounced (in my American English) [və.ˈlændə], that it refers to a type of porch, and that I’d only use it in reference to the sort of porch one finds in the southern part of the US or perhaps in some tropical country. Unless I was being ironic, I probably would not call my own back porch ‘the verandah’. I also know that *barf* is a verb that’s pronounced [bɑːf], that it means ‘vomit’, that it is intransitive (unless used with a particle like *up*) and that it is used only colloquially (I wouldn’t use it if I were describing the symptoms of a stomach flu to the doctor).

It is quite likely that in our mental lexicons we have entries that are only partial. We may know the pronunciation of a word, but not its meaning (e.g., I know how to pronounce *amortize*, but I’m not sure what it means). Or the opposite: for example, I know what the word *hegemony* means, but I don’t know if it’s pronounced with the stress on the first or second syllable. We may also have only partial knowledge of the meaning of a word. I know, for example, that a *distributor* is part of a car and that if you have to replace it, it’s a relatively expensive job, but I don’t know what a distributor looks like or what it does.

Each person’s mental lexicon is sure to contain things that are different from other people’s mental lexicons. One person may know lots of words for types of birds or flowers, another might know all the specialized vocabulary of sailing, and so on. Auto mechanics surely know more details of the meaning of the word *distributor* than I do. But our individual mental lexicons overlap enough that we speak the same language. In this section we will look in more detail at the contents of our mental lexicons, both what is stored and what is created by rules of word formation, and how our mental lexicons are organized.

1. A transitive verb is one that has both a subject and an object. An intransitive verb has only a subject.

2.3.1 How We Study the Mental Lexicon

Studying the mental lexicon has traditionally been the purview of psycholinguists. For the most part, theoretical linguists study language by studying linguistic data, and that is largely what we will do in this textbook. Psycholinguists and neurolinguists take a different path to the same goals. Psycholinguists who study the acquisition of words can do so by observing children. Psycholinguists can also do experiments using either children or adults as subjects. And increasingly, psycholinguists and neurolinguists are using brain-imaging techniques to study how words – simple and complex – are organized in the mind.

There are many ways of trying to tease out how we store, form, access, and process both simple and complex words. Some psycholinguists design experiments to ferret out properties of the mental lexicon. One very common type of experiment involves **lexical decision tasks** in which a subject has to decide whether a word they are exposed to is an existing word (like *packer*), a potential word (like *zaxer*), or a nonsense word (like *plurgle*). The time that it takes the subject to decide, called the **reaction time** or **response latency**, gives us a clue as to whether the subject treats a complex word as a whole or breaks it down into its component parts. Lexical decision experiments may be either unprimed or primed. In **primed** tasks the subject is exposed briefly to a word that is related to the word they will later react to. So, for example, they might briefly see the word *walked* and have to decide later whether *walk* or *walks* or *walker* is a word. The question is whether (and how) seeing the earlier word affects the time it takes to react to the later word. In unprimed tasks, experimenters just measure subjects' reaction times when asked to judge whether simple or complex words are real words (*walked* versus *zaxed* versus *plurgle*), without prior exposure to related words.

Psycholinguists also use a technique called **eye tracking** to study the way subjects process complex words. In an eye-tracking study, experimenters record where subjects' eyes land when they look at a complex word and whether or not their eyes backtrack in the course of reading a word or once they have moved on to a new word. These eye motions give us a clue as to how subjects process and access complex words.

Finally, in recent years psycholinguists have begun to study the organization of the mental lexicon with the use of different kinds of imaging techniques like **positron emission tomography (PET)**, which measure the level of blood flow to different parts of the brain, **electroencephalography (EEG)**, which measures electrical activity on the scalp, **magnetoencephalography (MEG)**, which measures magnetic activity in the brain, **functional magnetic resonance imaging (fMRI)**, a way of mapping activity in the brain over the course of time (Baayen 2014), and **event-related potentials (ERP)**, which measure electrical activity in neurons, and which, like fMRI studies, can track the brain's reaction to a linguistic stimulus over the course of time (Clahsen 2016). Brain-imaging techniques like these can help us to figure out what parts of the brain are activated as a listener processes or produces a simple or complex word.

Psycholinguists have used these techniques with both neurotypical subjects and non-neurotypical subjects. The latter might include subjects with various kinds of **aphasia**, that is, language impairment caused by trauma or illness. Or they might include individuals with genetic disorders that are known to affect linguistic abilities.

2.3.2 How Many Words?

Psycholinguists estimate that the average English-speaking six-year-old knows 10,000 words, and the average high-school graduate around 60,000 words. Paul Bloom describes how this estimate can be made (2000: 5):

Words are taken from a large unabridged dictionary, including only those words whose meanings cannot be guessed using principles of morphology or analogy ... Since it would take too long to test people on hundreds of thousands of words, a random sample is taken. The proportion of the sample that people know is used to generate an estimate of their overall vocabulary size, under the assumption that the size of the dictionary is a reasonable estimate of the size of the language as a whole. For example, if you use a dictionary with 500,000 words, and test people on a 500-word sample, you would determine the number of English words they know by taking the number that they got correct from this sample and multiplying by 1,000.

Children generally begin to produce their first words around the age of 1. Bloom calculates that between the ages of 1 and 18 we would have to learn approximately ten words every day to have a vocabulary of 60,000 words. It's worth pointing out, I think, that this figure just takes into account the words that we have stored (fully or partially) in our mental lexicon, and not the words – perhaps an infinite number of them – that we can create by using rules of word formation. We will return shortly to our knowledge of word formation rules and its relation to our mental lexicon. First, however, we will look more closely at how we acquire our mental lexicon.

2.3.3 The Acquisition of Lexical Knowledge

Psycholinguists have devised experiments to try to learn how children and adults are able to acquire words so easily. You might think that the learning of new words is a simple matter of association: someone points at something and says “flurge” and you learn that that something is called a *flurge*. This may be the way that we learn some words, but surely not the way we learn the majority of words in our mental lexicons. For one thing, not everything for which we have a word can be pointed at.

And even if someone points and says a word, it is often not clear from the context what exactly is being pointed out. Psycholinguists sometimes call this the **Gavagai problem**, following a scenario first discussed by the philosopher W. O. Quine. To summarize:

Picture yourself on a safari with a guide who does not speak English. All of a sudden, a large brown rabbit runs across a field some distance from you. The guide points and says “gavagai!” What does she mean?

One possibility is, of course, that she's giving you her word for 'rabbit'. But why couldn't she be saying something like "There goes a rabbit running across the field"? or perhaps "a brown one," or "Watch out!" or even "Those are really tasty!"? How do you know?

In other words, there may be so much going on in our immediate environment that an act of pointing while saying a word, phrase, or sentence will not determine clearly what speakers intend their utterances to refer to.

Besides, we are rarely in a situation in which someone is actively instructing us about the meanings of words; although parents may point to things in a picture book and name them for a child, or school children may be asked to memorize a list of vocabulary words, we learn most words without explicit instruction and seemingly with very little exposure. Although we do not know nearly enough about this subject, there are several things that we do know about how word learning occurs.

First, it is believed that both children and adults are able to do what the psycholinguist Susan Carey has called **fast mapping** (Carey 1978). Fast mapping is the ability to pick up new words on the basis of a few random exposures to them. In one experiment, Carey showed that children who were casually exposed to a new color name *chromium* during an unrelated activity (following instructions to pick up trays of various colors) were able to absorb the word and recall it even six weeks later. Experiments have shown that adults exhibit this fast mapping ability as well; while the ability to learn linguistic rules (say, of syntax or phonology) is thought to decline after puberty, the ability to learn new words remains robust.

Challenge

Here's an experiment you can try. Collect five or six objects. All but one of your objects should be familiar items (a bunch of keys, a mug, a pencil, etc.). One object, however, should be something odd and not familiar to many people. Put all your objects on a tray, and ask your subject (anyone outside your class will do) to point out the *zorch*. Observe what your subject does. Now take away the unfamiliar object, leaving only the familiar objects, and ask a different subject to point out the *plitz*. Again, observe closely what the subject does.

Psycholinguists have proposed a number of other strategies that both children and adults seem to use in learning new words.² One might be called the **Lexical Contrast Principle**. For example, in an experiment similar to yours, children were asked to point to the *zorch* (or some other made-up word), and what they invariably did was to point out the unfamiliar object. According to the Lexical Contrast Principle, the language

2. See Bloom (2000) for an extensive discussion of this subject.

learner will always assume that a new word refers to something that does not already have a name.

A second word learning strategy might be called the **Whole Object Principle**. In the experimental condition described above, when subjects are presented with the word *zorch* and an unnamed object, they will assume the whole unnamed object to be a *zorch*. They will not assume that *zorch* refers to a part of the object, to its color or shape, or to a superordinate category of objects to which it might belong.

A related strategy might be dubbed the **Mutual Exclusivity Principle**. In the second experiment above, there are only familiar objects for which subjects already have names. When asked to point out the *plitz*, experimental subjects typically do one of two things: they might first look around the room for something else that might be called a *plitz*, or they might assume that the word *plitz* refers to a part of one of the familiar objects or a special type of one of them. Subjects, in other words, will assume that if an object already has a word for it, the word *plitz* cannot be synonymous with those words.

These experiments are of course not just hypothetical. Paul Bloom, Susan Carey, and many other psycholinguists have conducted them both with children of various ages and with adults, and have obtained the results described above. What is perhaps most astonishing about their results is that their experimental subjects often remember the words they've been exposed to when they are retested weeks after the original experiment. But maybe we should not be surprised by this: how otherwise could we have learned 60,000 words by the time we're 18?

Children not only learn simple words, but – as we'll see in the chapters to come – they learn the rules that allow us to create and understand new complex words. English-speaking children acquire some rules of word formation quite early on; for example, [Berman \(2009\)](#) notes that English-speaking children are known to produce compound words (roughly, words made up of two or more independent words) as early as 18–24 months, and 3-year-olds acquiring English can come up with novel compound words like *wagon boy* when asked what they would call a boy who pulls wagons. The same appears to be true of children acquiring other Germanic languages like German, Dutch, and Swedish, but less so for languages like Hebrew and French, where compounding is not used nearly as much in adult language as in the language of adults speaking Germanic languages.

Children also use prefixed and suffixed words early on, but as [Clark \(2014\)](#) points out, we only know when they have learned the rules for forming prefixed and suffixed words when we see that they can produce new words or say what new complex words mean; in other words, a child may know the word *baker* as a whole chunk without knowing that it's made up of two pieces, but when a child can tell us that someone who *zaxes* is a *zaxer* they show us that they know what the suffix *-er* means and how it can be put together with a verb to form a new noun. [Clark \(2014\)](#) indicates that some suffixes like *-er* and prefixes like *un-* are acquired very early by English-speaking children, emerging at around 3

years of age but becoming better established by 5. According to Clark, children acquiring French start using prefixes and suffixes as early as 2 years old. It seems that because French relies on prefixes and suffixes more than compounding, these elements of word formation begin to emerge earlier in French-speaking children.

Novel Uses of Complex Words in Child Language (Clark 2014: 431–2)

D: (2:5.26, as he reached across the kitchen counter): *I'm a big reacher.*

D: (3:5.27): *did you know Gabriel is a stealer? sometimes he steals Judy's things.*

D: (3:6.25 ...): *These are my pantsies. D'you know what I call a cat? ... A cattie. D'you know what I call a bear? A bearie! D'you know what I call a dog? A doggie!*

2.3.4 The Organization of the Mental Lexicon: Storage versus Rules

Although linguists like to describe our knowledge of words as a mental lexicon, we know that the mental lexicon is not organized alphabetically like a dictionary. Rather, it is a complex web composed of stored items (morphemes, words, idiomatic phrases) that may be related to each other by the sounds that form them and by their meanings. Most linguists also believe that along with these stored items we also have rules of some sort that allow us to combine morphemes in different ways, a subject into which we will delve in the chapters to come. The question that we must first ask is what exactly our mental lexicons contain and how that content is organized. Certainly we must have simplex words stored in our lexicons, and our ability to form new complex words suggests that we have rules stored, but how do we know? And what about complex words that we have already encountered? Do we store them as unanalyzed wholes? Or do we use word formation rules to recreate a familiar inflected or derived word every time we use it or break it down every time we hear it? This has sometimes been called the 'words *versus* rules' debate. All of the experimental techniques described in [Section 2.3.1](#) have been deployed at one time or another in this debate. What we have learned is that the picture is complicated and experimental results frequently do not agree. Nevertheless there seems to be substantial psycholinguistic evidence that the organization of the mental lexicon involves an intricate combination of storage of both simple and complex words and creation of complex words via rule ([Gagne and Spalding 2019: 552](#)). It also appears that storage *versus* rules is not necessarily an either/or matter. Apparently whether we access a complex word like *packer* whole or create it anew using our rules depends on many factors including our familiarity with the complex word (i.e., the frequency with which we have encountered it before), the frequency of the root or

base on which the complex word is formed (in this case, *pack*), the number and frequency of the family of words containing the same root or the same prefix or suffix (e.g., *packing*, *package*, *pack rat*, and so on), and a variety of other factors. Psycholinguists have by no means come to an agreement on how our mental lexicons are arranged, but this continues to be an active area of research in morphology. We will illustrate some of the complexities of the debate by looking at the past tense of verbs in English, a subject on which there has been ongoing debate. Let's first take a quick look at how we form past tense verbs in English.

At this point, if I asked you how to form the past tense of a verb in English, you would probably say that you usually add an *-ed*. And then you might point out that there are a number of verbs that have irregular past tenses like *sing-sang*, *tell-told*, *win-won*, *fly-flew*, and the like. While it is true that in writing we add an *-ed* to form the past tense of a verb, in terms of spoken speech, the situation is a bit more complicated. Consider the next Challenge.

Challenge

Consider how you *pronounce* the past tenses of these verbs:

1. rap, tack, laugh, sheath, pass, lurch
2. pat, prod
3. rob, rove, bathe, buzz, rouge, judge, warm,
warn, bang, roar, rule, tango

Transcribe the past tenses of these words in the International Phonetic Alphabet and observe how they differ.

You pronounce the past tenses of the first set of words in the Challenge box with a [t] sound, in the second with a sound like [əd], and the third with a [d] sound.

We do not choose the pronunciation of the past tense at random. Rather, the choice of which of the three endings to use depends on the final sound of the verb. Those words that are pronounced with final [t] or [d] sounds – those in the second list – get the [əd] pronunciation. The words that end in voiceless (with the exception of [t]) sounds get the [t] pronunciation. And all the rest get the [d] pronunciation. As for irregular forms like *sang* and *flew*, we can assume for now that English speakers simply learn them as exceptions.

We know that speakers of English have an unconscious knowledge of the regular past tense rule because we can automatically create the past tense of novel verbs. For example, if I coin a verb *blick*, you know that the past tense morpheme is pronounced [t]. Similarly, the novel verb *flurd* will have the past tense [əd], and the verb *zove* will be made past tense with [d]. We can even form the past tense of verbs that contain final sounds that do not occur at all in English, and when we do, we still follow

the rule. For example, if we imagine that there are many composers imitating the style of Johann Sebastian Bach, and we coin the verb *to bach* to denote the action of imitating Bach, we will automatically form the past tense with the past tense variant pronounced [t], because the final sound of Bach is [x], a voiceless velar fricative. The important point here is that when we hear this sound at the end of a verb we know (unconsciously) that it's voiceless, and apply the past tense rule to it in the usual way.

Now that we know something about the English past tense, we can return to the question of how the mental lexicon is organized, and specifically the “words *versus* rules” debate (Pinker 1999). A “words *versus* rules” model of the mental lexicon hypothesizes that speakers of English use the past tense rule every time they create the past tense of a regular verb, but simply store the past tense forms of irregular verbs whole and unanalyzed in their mental lexicons. We will see that the evidence from acquisition, experimentation, and studies of non-neurotypical populations sometimes yields conflicting evidence regarding this hypothesis.

2.3.5 Evidence from Language Acquisition

It has been known, ever since the pioneering work of the psycholinguist Roger Brown (1973) that children acquiring English exhibit a general pattern in which they first correctly use a few highly frequent irregular past tenses (for example, *came*, *brought*). Following this, there comes a period in which children start to use regular past tenses (*walked*, *played*). At this point we begin to observe what is called ‘over-regularization’: children who previously used *came* and *brought* appropriately might start saying *comed* and *bringed* instead, suggesting that they have learned a rule for creating the past tense. Only somewhat later do children go back to using the irregular forms appropriately. One common interpretation of this pattern is that children first store past tenses, then learn rules, and finally sort out the rule-governed cases from the stored ones.

In recent years, psycholinguists have looked at child acquisition data in more detail and have observed interesting statistical patterns in the past tenses that children use. For example, Lignos and Yang (2016) report that although children frequently over-regularize the regular past tense, they almost never over-irregularize, that is, create an irregular past tense for a verb with a regular past tense. Even more interestingly, Lignos and Yang (2016: 772) use statistical patterns in child language to argue that some irregular forms may be rule-governed rather than simply stored. They first observe that “irregular past tense forms that are more frequently used in the input tend to be acquired more accurately by children.” Interestingly, frequency for individual verbs does not seem to matter as much as frequency of a whole irregular pattern. For example, *bring*, *catch*, *fight*, and several other verbs all have past tenses that end in /ɔt/. It is the total frequency of all of these verbs that correlates with the accuracy of past-tense learning in children, rather than the frequencies of the individual verbs, which suggests that there is a subrule that children learn in addition to the main past tense rule.

2.3.6 Evidence from Aphasia

Studies of individuals with aphasia have also contributed to the “words versus rules” debate. For example, studies by [Badecker and Caramazza \(1999\)](#) and [Ullman et al. \(2005\)](#) with different kinds of aphasic subjects offer some support for the hypothesis that irregulars are stored and regulars generated by rules. Specifically, these studies suggest that irregular past tenses are stored in one area of the brain and rules creating regular past tenses are located elsewhere.

Some aphasics display a form of aphasia called **agrammatism**. Also called non-fluent aphasics, these individuals speak haltingly and with effort and have difficulty in producing or processing function words in sentences. However, they can still produce and understand content words. Experiments seem to show that when asked to read or produce the past tenses of different kinds of verbs, non-fluent aphasics have more trouble with regular past tenses than with irregular ones. These subjects tend to have damage to frontal areas of the left hemispheres of the brain that have been associated with rule-based linguistic activity. In contrast, so-called fluent aphasics produce syntactically complex sentences but have trouble accessing content words. When asked to read or produce the past tenses of regular and irregular verbs, they perform more accurately for regulars than for irregulars. And their linguistic deficits seem to be associated with damage to the posterior part of the left hemisphere. Linguists like [Ullman et al. \(2005\)](#) have suggested that this evidence supports the idea that morphological rules are associated with a different part of the brain than items that are stored. For non-fluent aphasics, the rule is unavailable, presumably because the part of the brain has been damaged that apparently allows us to apply morphological rules, but the irregular forms are still accessible from an undamaged part of the brain. For fluent aphasics, the irregular forms have been lost because the part of the brain that apparently allows access to stored forms has been damaged, but the regular rule is still intact. This finding is suggestive, but as [McWhinney \(2005\)](#) rightly points out, it must be taken as tentative for two reasons: first, because the number of subjects studied is very small; and second, because currently available studies do not absolutely rule out the possibility that frequent regular past tenses can be stored for some non-fluent aphasics.

2.3.7 Evidence from Genetic Disorders

Studies of subjects with genetic disorders yield a similar picture. Psycholinguists have studied two disorders in particular – **Specific Language Impairment (SLI)** (also called **Developmental Language Disorder**, [Bishop 2017](#)) and **Williams Syndrome** – that affect language in different ways. Individuals with SLI are generally of normal intelligence and have no hearing impairment. But they are slow to develop and understand language, and their speech is characterized by the omission of various inflectional morphemes. Individuals with Williams Syndrome have a genetic disorder linked to various heart problems, elevated levels of calcium in their blood, and a characteristic appearance (short stature,

an upturned nose, a long neck, among other things). Their language and social skills are in the normal range, but in other respects such as motor control and spatial perception they display mild or moderate developmental delay.

What is significant for our purposes is that these disorders are also thought to provide evidence for the organization of our mental lexicon. Individuals with SLI find it difficult to create the past tenses of novel verbs, and often fail to inflect unfamiliar regular verbs correctly; they have less difficulty with irregular verbs, though. In spontaneous speech, they may leave the regular past tense off verbs (Redmond and Rice 2001; Marshall and van der Lely 2012). In contrast, individuals with Williams Syndrome speak fluently and produce sentences with correct regular past tenses, but have more trouble with irregular ones; indeed they seem to use regular past tense marking even where control subjects or individuals with SLI would not, for example, overgeneralizing the regular *-ed* ending on irregular verbs (e.g., *falled*) (Clahsen, Ring, and Temple 2004). Assuming that the genetic anomalies associated with these disorders affect different parts of the brain, we can explain this pattern of behavior.

2.3.8 Evidence from Imaging Studies

Early imaging studies of normal subjects, done with PET scans, seemed to yield results that support the ‘words and rules’ hypothesis. Jaeger et al. (1996), for example, reported that there are parts of the brain that are activated when subjects are asked to read regularly inflected past tenses that are distinct from those activated in reading or producing irregular past tenses. More recent ERP studies that tested both perception (Munte et al. 1999) and silent production of regular and irregular past tense forms (Budd et al. 2013) found evidence that subjects showed different reactions to regular and irregular past tenses.³ Such results are again argued to support the ‘words and rules’ hypothesis.

But these results are not uncontroversial: in another study, Desai et al. (2006) point out that imaging studies do not agree on which parts of the brain are associated with regulars and irregulars, which is somewhat troubling. And in a study using MEG imaging Stockall and Marantz (2006) argue against the ‘words and rules’ hypothesis, suggesting that subjects do not access irregular past tenses as unanalyzed wholes in their mental lexicons, but rather activate both the irregular form and the root of the verb – suggesting that some analysis is going on early in the process of recognizing the past tense. It seems safe to say that this is an extremely important and interesting area of study, but that we are not able to draw firm conclusions yet about the organization of the mental lexicon.

3. Clahsen (2016) reports that regular past tenses showed a reduced N400 component in the ERP in comparison to irregular past tenses. N400 is “a negative waveform that peaks at around 400 ms after stimulus onset ...” (Clahsen 2016: 795).

2.3.9 Reprise: Is It Really a Word?

We have spent some time in this chapter contrasting the dictionary with the mental lexicon in order to understand the question “Is xyz really a word?” We are now in a position to understand this question better, at least from the point of view of morphologists. Most morphologists would say that xyz is a word if it can be formed by the rules of word formation in a particular language. So words like *wordhood* or *re-reprise* that you might never have seen before you read this chapter really are words, even though you won’t find them in any dictionary. They are words because they follow the rules of English word formation. It is the rules of word formation that we know that most distinguish our mental lexicon from the dictionary. The dictionary does not need to list all the words that we know or that we could create, because once we know word formation rules we can produce and understand potentially infinite numbers of new words from the morphemes available to us. The remainder of this book will be an attempt to work out in some detail what those rules are.

2.4 More about Dictionaries

In [Section 2.2](#) we considered all the reasons why morphologists don’t look upon dictionaries as the ultimate arbiters of ‘wordhood’ in English, or indeed in any language. You may not need more convincing of this issue, but for those of you who have a fondness for dictionaries (most morphologists do!), it’s worth knowing something about how dictionaries have developed. I’ll again concentrate on English here, as our common language, but the history of dictionary-making for other languages can be equally fascinating.

Let’s start with a thought experiment. Look at the next Challenge.

Challenge

Suppose that a great catastrophe has occurred and every single written or online dictionary has disappeared from the face of the earth. You and your classmates have survived the catastrophe (perhaps in a hidden concrete tunnel beneath the building in which you are now sitting), and have been delegated the task by other survivors of creating the first post-catastrophe dictionary of your language.

How would you start?

Your first instinct would probably be to make a list of words that you would need to define. Assuming that there were no surviving books to use as dictionary-fodder, a good way to begin would be by thinking of categories, and listing everything you could in each one. After you’ve listed all the animals, plants, and types of furniture you could think of,

you'd come up with a list of hairstyles (*crewcut, bob, beehive, bun, buzz cut, duck's ass, cornrows, mullet ...*) and condiments (*ketchup, soy sauce, mustard, horseradish, wasabi, sambal oelek ...*), and so on, and eventually you'd come to articles (*a, the, this, that ...*), prepositions (*in, on, above, during, for ...*) and the other small words that form the grammatical glue that holds sentences together.

But along the way, you'd discover a number of problems. First, you'd have a suspicion that you'd be forgetting things (what, for example, was the name for that women's hairstyle that was the rage in the seventies?). Second, you and your classmates would get into constant arguments over this word or that: is it worth putting the word *mullet* in the dictionary as the name of a hairstyle? Wasn't that slang? Does slang go in the dictionary? What IS slang, anyway? Is it too vulgar to put *duck's ass* in the dictionary as a name of a 1950s hairstyle? What about really raunchy words? Is *sambal oelek* a word for a condiment in English, or is it just something we've borrowed from another language (what other language, though?)?

What this thought experiment does is to put you in the shoes of a lexicographer. In reality, it's been centuries since lexicographers have had to start from scratch in creating a dictionary – and perhaps they've never really done so. As the lexicographer Sidney Landau has said about the tradition of dictionary-making in English (2001: 43), "The history of English lexicography usually consists of a recital of successive and often successful acts of piracy." For years and years, each succeeding dictionary-maker has consulted already existing dictionaries to come up with a base list of words, often adding new ones and sometimes deleting words for various reasons. But at least at first, lexicographers did have to decide one by one on each of the English words to include. Of course, there were manuscripts and books available to suggest words that needed to be included, and in fact, the earliest English lexicographers did rely on the words they found in books as the material from which they built their dictionaries.

2.4.1 Early Dictionaries

It was not until the early seventeenth century that anything we would recognize as a monolingual dictionary could be found for the English language. Dictionaries or glossaries for translating Latin to English date from a century or so earlier, and in the sixteenth century lists of so-called hard words could be found for English, explaining words which largely had been adapted from Latin. The first real dictionary of English is generally acknowledged to be Robert Cawdrey's (1604) *A Table Alphabeticall of Hard Words*. The tradition of lexicographical piracy goes at least as far back as Cawdrey, who is said to have used an available Latin–English dictionary of his day to help come up with the words to define. The first dictionaries going beyond the tradition of defining only 'hard' words to include ordinary, everyday words began to appear in the early eighteenth century; Landau (2001: 52) cites John Kersey's (1702) *A New English*

Dictionary as the earliest of these, followed by Nathaniel Bailey's (1721) *An Universal Etymological English Dictionary*.

2.4.2 Johnson's Dictionary

A more significant milestone in the history of English lexicography for our purposes was Samuel Johnson's *Dictionary of the English Language*, published in 1755. It contains more than 42,000 entries – even then, only a small fraction of English vocabulary – and took seven years to write, an astonishing feat for a single individual. Johnson's dictionary was not only the most comprehensive English dictionary of his time, but it was also among the first dictionaries to include illustrative quotations on a large scale.

What is most interesting for our purposes, though, are the idiosyncrasies of Johnson's dictionary: what he included, what he left out, and how he defined various words in odd ways. Henry Hitchings (2005: 110) notes that:

dictionaries are fraught with submerged ideas, narratives and histories. Johnson's is no exception. It offers no overarching system of knowledge, but it is a literary anthology, a compendium of quotable nuggets, and a mine of information – some trivial, some considerable – on subjects as diverse as heraldry and hunting, rhetoric and pharmacy, oracles and literary style, the zodiac and magic, law and mathematics, ignorance and politics, the art of conversation and the benefits of reading.

Johnson's dictionary, in other words, contains a lot about Johnson himself – both his interests and his prejudices. It was quite a comprehensive dictionary in its time. But Hitchings notes that Johnson still left out entries for such words as *ultimatum*, *irritable*, *zinc*, *engineering*, *athlete*, and *annulment*, even though he actually used some of those words in his definitions. On the other hand, he included such words as *ariolation*, *clancular*, *deuteroscopy*, and *impossibility*, which even the nineteenth-century American lexicographer Noah Webster considered dubious. And it has often been pointed out that some of Johnson's definitions were odd, unhelpful, and occasionally downright wrong.

For example, the word *urim* is used by Milton, and therefore Johnson judged it important enough to be included even though he was unable to discern the meaning of the word from its literary context. Similarly, *trolmydames* is used in Shakespeare, and therefore it merited inclusion for Johnson – although, again, he had no idea what it meant. And what can we say about the definitions for *network* and *worm*? If you don't already know what they mean, you won't be enlightened by Johnson's definitions!

As Hitchings implies in the passage quoted above, we can learn a lot about Johnson's idiosyncrasies from his dictionary. For example, we can tell from Johnson's entry for *pastern* that he had no particular knowledge of or interest in horses: he defines the *pastern* as the knee of a horse. People who are interested in horses know that a *pastern* is part

Some
Johnsonian
Definitions:

urim: Urim and thummim were something in Aaron's breast-plate; but what, critics and commentators are by no means agreed.

trolmydames: [Of this word, I know not the meaning.]

worm (v.): To deprive a dog of something, nobody knows what, under his tongue, which is said to prevent him, nobody knows why, from running mad.

network: Anything reticulated or decussated, at equal distances, with interstices between the intersections.

pastern: The knee of an horse.

of a horse's foot. Similarly, as Hitchings points out, Johnson apparently had no interest in music. His definitions for a number of stringed instruments (*viola*, *lute*, *guitar*) are precisely the same: "a stringed instrument." Furthermore, the definition of *violin* suffers from the cardinal lexicographical sin of circularity: the entry for *violin* sends one to the entry for *fiddle*, which in turn sends one back to *violin*.

These examples are not intended to imply that Johnson's dictionary was incompetent – far from it, it was an amazing achievement for one man working alone for seven years. Much of it still holds up to twenty-first century scrutiny. For every entry that is obscure, weird, or unhelpful, there are a hundred that are brilliant and insightful. I devote this much attention to its deficiencies merely to point out that dictionaries are fallible, and often reflect the foibles of their makers.

2.4.3 Webster's Dictionary

Johnson's dictionary was followed in 1828 by Noah Webster's dictionary – billed as the first American dictionary. Webster's agenda in writing his dictionary was at least partly political; through the dictionary he sought to establish American English as a national language. His dictionary included not only new words but also new meanings that had developed for old words in the context of American life, for example, words relevant to the newly minted form of democracy, such as *congress* and *senate*. Webster is also credited with promoting the spelling differences which even today distinguish American from British English – *color* instead of *colour*, *center* instead of *centre*, *tire* instead of *tyre*, and so on.

Webster was not particularly skilled at **etymology** (the study of where words come from); Baugh and Cable (2013: 356) suggest that his sense of nationalism caused him to ignore advances in historical and comparative linguistics that were taking place in Europe at that time. However, his definitions are excellent. Not surprisingly, though, some definitions in Webster's dictionary are pirated directly from Johnson. Note, however, that not all of Webster's contemporaries shared his desire to distinguish American English from British English. Joseph Worcester, for example, published his own *Comprehensive Pronouncing and Explanatory Dictionary* in 1830, in which he took a far more conservative approach to Americanisms and spelling.

2.4.4 The Oxford English Dictionary

By the mid-nineteenth century, members of the English Philological Society had come to feel that Johnson's dictionary was inadequate. As we saw above, Johnson had missed many words, and even if he had not, over the course of a century many new words are added to a language and many old words come to be used in new ways. After much deliberation and a number of false starts, the Philological Society chose James Murray, a Scottish schoolmaster, to edit the *New English Dictionary*. Oxford University Press contracted to publish it, and by 1895 it had come to be known as the *Oxford English Dictionary*. Murray began work on the dictionary in 1879, hoping to finish it within 10 years. But it would be almost

50 years before the first edition of the dictionary was finished, during which time three more editors were added, and Murray himself died.

The *OED* took so long to compile because the goals of its originators were so ambitious. Murray and his colleagues sought to create a dictionary that would not only give current meanings of words, but also trace those words back as far into the history of English as they could, taking note of all the spelling variants and meaning changes along the way. Following Johnson's dictionary, all senses of words would be illustrated with quotations from literary works. Words that were already archaic or obsolete by the late nineteenth century would still be included, as long as they had not died out before 1250 CE. The dictionary was to be comprehensive in both breadth and depth, a task which turned out to be far more challenging than anyone in 1879 could have anticipated. The first edition of the *OED* ran to 10 large volumes and contained almost a quarter of a million main entries. By the time the last volume was finished, the early volumes were already obsolete; one supplement was added in 1933, and a second one in 1972. A second edition of 20 volumes was issued in 1989, incorporating all of the supplements into the original volume. Today, work continues on the third edition, with segments issued online on a quarterly basis, as they are finished. Since the first edition, the *OED* has grown to include more than half a million entries; in its online form, size and space are no longer as much of a concern as they once were.

James Murray was well aware both of the weight his lexicographical decisions carried and of his potential fallibility in making those decisions – after all, most people do look to the dictionary to determine whether *xyz* really is a word. Perhaps Murray put it best when he noted in the Introduction to the first edition that:

The Vocabulary of a widely-diffused and highly-cultivated living language is not a fixed quantity circumscribed by definite limits. That vast aggregate of words and phrases which constitutes the Vocabulary of English-speaking men presents, to the mind that endeavours to grasp it as a definite whole, the aspect of one of those nebulous masses familiar to the astronomer, in which a clear and unmistakable nucleus shades off on all sides, through zones of decreasing brightness to a dim marginal film that seems to end nowhere, but to lose itself imperceptibly in the surrounding darkness.

In other words, it's impossible to pin down the vocabulary of English (and we might add, any other language). Murray illustrates his point with a diagram (Figure 2.1), reproduced from the Introduction to the first edition of the dictionary. His idea is that there is a core of words whose place in the dictionary nobody would dispute, encompassing what he called "common," "literary," and "colloquial" words. Common words are words that occur in all registers of English, like *mother*, *dog*, *walk*, *apologetic*, *wiggle*, *if*, *and*, *to*, *in*, *that*, and so on. Literary words are words that we might recognize when we read, but would not necessarily use in daily conversation, words, for example, like *omnipotent*, *notwithstanding*, *heretical*, *hegemony*, and *ambulatory*. And also among the core

Of Obscure
Meaning in the
OED

These entries for
smazky and *val-
dunk* are among
the 87 entries that
the *OED* design-
ates as "meaning
obscure" or "of
obscure meaning."
Did they deserve
to be in the *OED*?
You decide!

smazky, a. Obs.
(Meaning obscure.)

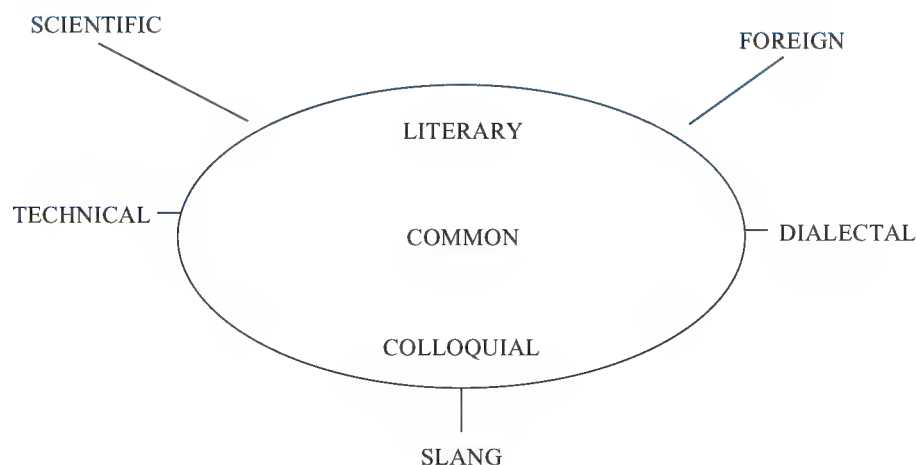
1599 MIDDLETON
Micro-cynicon
A5, Auant, ... Ile
anger thee inough,
And fold thy fry-
eyes in thy smazkie
snufe.

val-dunk Obs.
(Meaning obscure.)

1631
R. BRAITHWAIT
Whimzies,
Winesoaker 102,
By this time his
cause is heard, and
now this val-dunke
growne rampant-
drunke, would
fight if hee knew
how.

FIGURE 2.1:

Murray's view of English vocabulary. Reproduced with permission of Oxford University Press, James A. H. Murray et al. 1888. *A New English Dictionary on Historical Principles*, p. xvii



words would be colloquial words, ones that we use frequently in spoken language, but far less frequently in written or formal language, for example, *grubby*, *pooch*, and *mad* (in the sense of ‘angry’). But there is no clear dividing line between these words and words which are perhaps too technical or scientifically specialized (*circumfix*, *triptan*), not quite assimilated enough into English (*tchachka*, *sambal oelek*), too bound to a specific dialect (*frappe*, *black ice*), or too informal, impermanent, or bound too narrowly to a particular time or a particular segment of society (*groovy*, *homie*). Deciding which of these uncommon words merit inclusion in the dictionary is a judgment call, often based more on practical considerations – the size of the dictionary, its intended audience – than on strict linguistic principles. All lexicographers face this conundrum, and each one makes a slightly different decision.

The *OED* is certainly the gold standard for English dictionaries today. Nevertheless, it has its own idiosyncrasies. For example, it contains a number of nonce words, words that are attested only once. Indeed, there are quite a few nonce words that the *OED* includes, even though it is unable to define them. A search, using the key words “meaning obscure” and “of obscure meaning,” turns up 87 words so labeled, including *smazky*, *squirglit-ing*, *val-dunk*, *vezon*, *uncape*, and *umbershoot*. Each bears an entry illustrated with one quotation, which unfortunately does not illuminate the word’s meaning. Nevertheless, these and 81 others like them made it into the *OED*!

2.4.5 Modern Dictionaries

Today, there are dozens of dictionaries available for English – unabridged dictionaries, college dictionaries, children’s dictionaries, specialized dictionaries of music or architecture, an official Scrabble dictionary, not to mention online dictionaries in many varieties. Each one of these is edited by a team of individuals who make the judgment call whether xyz deserves to be in the dictionary. The decision is made on a number of grounds:

- the size of the dictionary, which determines the number of words it can hold;

- the intended audience of the dictionary (adults, children, language learners, etc.);
- whether a word has a sufficiently broad base of usage;
- whether it's likely to last;
- whether it's too specialized or technical for the intended audience;
- for a word borrowed from another language, whether it's assimilated enough to be considered part of English.

With respect to size, the number of words and the depth of entries (whether etymologies and illustrative quotes are included, for example) in print dictionaries are determined by the number of pages and the font size of the print used. Online dictionaries do not have the sort of space constraints that print dictionaries do. As for audience, a dictionary intended for college-age adults will probably have more learned and technical words than a dictionary for children. On the other hand, words that a native speaker is unlikely to need defined might be more of a focus in a dictionary for English language learners; the meanings of prepositions and their idiomatic uses come to mind here. The type of dictionary also determines how broad a base of usage a word needs to have in order to be included. Dictionaries of slang, dialect, or specialized fields obviously contain more narrowly used words than general dictionaries do (although you might be surprised at how much slang and technical terminology can be found in general dictionaries).

Perhaps the trickiest issue is how long a word has to have been around to merit inclusion in the dictionary. These days, words can appear in dictionaries fairly quickly, especially in online dictionaries. The *OED* already has entries for the words *flexitarian*, *hashtag*, and *selfie*; the first citation for the last of these was as far back as 2002. The word *bounce-backability* – allegedly coined by British sportscaster Iain Dowie in 2004 (Hohenhaus 2006) – already had a draft *OED* entry by June, 2006 (although, interestingly, the *OED* has traced the word as far back as 1961!).

And there is more to consider in deciding whether a word goes into the dictionary. Take, for example, words formed with various prefixes and suffixes. If *happy* is in the dictionary (as it certainly would be), do we need to have an entry as well for *happiness*? Similarly, if *sad* has an entry, do we need *sadness*? If our audience is a learner of English, perhaps yes, but for native speakers who know intuitively how the suffix *-ness* is used, is there any need for these extra entries? Interestingly, dictionaries are often quite inconsistent on how many and which derivatives with particular suffixes get entries. The online *OED* has entries for *redness*, *blueness*, *pinkness*, *greenness*, and *yellowness*, but not *orangeness*. The word *purpleness* is used in the definition of the word *purplely*, but does not have its own entry. And not surprisingly, there is no entry for *mauveness* or *beigeness*. What is more surprising is that there are so many entries for color words with the suffix *-ness* attached.

Certainly, if a word derived with a prefix or suffix takes on an idiosyncratic or **lexicalized** meaning, the dictionary needs to include it.

Take, for example, the word *transmission*, which can have the transparent meaning ‘the act of transmitting’ but probably more often is used to denote a part of a car. This second meaning probably deserves to be in the dictionary. But is it necessary to include all derived words whose meanings are perfectly clear from the meaning of the base plus the meaning of the affix? Probably not.

Until the last decade of the twentieth century lexicographers made their decisions by reading materials of all sorts, and in more recent decades by listening to radio, TV, and talk in general. Potential entries would be recorded with their context on small slips of paper. These slips would then be filed, and when a critical mass of usages accumulated for a word, it might be considered for entry in the dictionary. These days lexicographers are aided by *corpora* (singular *corpus*), large computerized databases that can be searched for words in the context of their use, and by the internet, which might be viewed as a vast corpus. Indeed the rise (and sometimes fall) of a new word can be traced by searching for its use on the internet. We will see in the [next chapter](#) that corpora can also be invaluable in the study of word formation rules.

Perhaps the most interesting recent development in lexicography is the rise of Wiktionary – an online collaborative dictionary created not by professional lexicographers, but by users themselves (www.wiktionary.org). In the instructions for submitting entries, Wiktionary asks that words be attested, by which it means they must be in widespread use, available in well-known works or refereed publications, and used at least three times in at least three sources over more than a year. It does, however, have a category of what it calls ‘protologisms’ for “terms defined in the hopes that they will be used, but which are not actually in wide use.”

One final note about the vagaries of dictionaries. Lest you think that lexicographers are humorless (“harmless drudges” as Johnson calls them in his dictionary), let’s consider the issue of mountweazels mentioned briefly above. As Henry Alford reveals in the August 29, 2005 issue of *The New Yorker*, the editors of the *New Oxford American Dictionary* (2001) planted the non-existent word *esquivalience* (defined as “the willful avoidance of one’s official responsibilities”) among the entries for the letter “e” to catch potential dictionary pirates. Such false words are called ‘mountweazels’, from the fictitious entry for Lillian Virginia Mountweazel in the *New Columbia Encyclopedia*.⁴ What is most interesting for our purposes is that once these fake words have been coined, they take on lives of their own. As of July 2020, there were 10,900 hits for *esquivalience* and 49,200 for *mountweazel* on Google, leading me to wonder whether these fakes have now become real words.

4. According to the fictitious entry in the *New Columbia Encyclopedia*, Lillian Virginia Mountweazel lived from 1942 to 1973. A fountain designer and photographer, she was supposedly well known for taking pictures of rural American mailboxes. She died tragically in an explosion while she was on assignment for *Combustibles* magazine.

2.4.6 And Other Languages

I have concentrated here on the history of dictionary-making in English, but the same points might be made with respect to dictionaries of French, Italian, Russian, Chinese, or Central Alaskan Yup'ik. All dictionaries are products of individuals and all display the choices and idiosyncrasies of those individuals in some way or another.

Dictionaries of other languages might be organized quite differently from those of the Indo-European languages that we are most familiar with, however. For example, dictionaries of Mandarin Chinese are not alphabetized in the way that dictionaries of English and French are, because Chinese is not written in the Roman alphabet. Instead, the writing system (or **orthography**) of Chinese is **logographic** or word-based. Each word in Chinese is represented by a single character (or sometimes a combination of two characters). When you look up a word in a Chinese dictionary, you need to know how many strokes or lines make up that character. Dictionaries are organized from those characters made up of the fewest strokes to those containing the most strokes.

Dictionaries of other languages might include many fewer complex words than English dictionaries typically do. For example, if a language has very regular rules of word formation such that both the form and the resulting meaning of a complex word are perfectly predictable, the dictionary will have no need to list all complex words in separate entries. All it needs to do is list individual morphemes with their meanings (and perhaps some indication of how they combine). But the less predictable the form and meaning of complex words are, the greater the need to put them in the dictionary.

Summary

In this chapter we have been concerned with the question of what constitutes a word. We have contrasted dictionaries with the mental lexicon. Dictionaries are written constructs that record words, along with their pronunciations, meanings, etymologies, and perhaps examples of use. On the one hand, they do not and cannot contain everything that native speakers would recognize as words of their language – dictionaries have no need to record regularly inflected forms of words and words derived by very active rules of word formation, for example. On the other hand, dictionaries may include items that perhaps don't deserve to be considered real words – for example, nonce words that are undefinable, or artificially created words put in to check for copyright violations.

Our mental lexicons are something different, however. High-school educated adults may have vocabularies of 60,000 words. We acquire these words rapidly, and sometimes our mental representations are sketchy or incomplete. The evidence we have looked at from aphasia and genetic disorders, as well as studies using

various imaging techniques, allows us to begin to develop a picture of how these vast numbers of words are organized in our minds, although the picture is still quite unclear. Unlike dictionaries that list words alphabetically, our mental lexicon is organized as a complex web of entries that are linked in various ways, along with a system of rules for combining listed forms. It appears that entries and rules are at least to some extent wired into different parts of the brain.

Exercises

1. Go to the *OED Online* website and search for words that are in the dictionary but have no known definition. To do this, click on Advanced Search (look towards the top of the *OED* home page), and type into the first open box "meaning obscure" or "of obscure meaning." Then choose three words and read through their entries. Do you think the *OED* was justified in including these words? If so, why? If not, why not?
2. Make a list of five words that you consider to be *slang*. Now look them up in your dictionary (you may use any dictionary at hand, whether print or online). First note whether or not you find them. If you do, is the dictionary definition the one that you had in mind? Does your dictionary list them as slang? If not, speculate on why they might not be listed as slang.
3. Make a list of at least ten words that come to mind that end in the suffix *-less*. Look these words up in a dictionary (you may use a standard college desk dictionary like the *American Heritage Dictionary* or you may use the online *OED*). How many of your words are in the dictionary? Is there any pattern that you can discern with respect to the words that are listed, as opposed to the words that are not?
4. Visit the Word Spy website (www.wordspy.com). Look at the list of new words and decide which ones, if any, are part of your own mental lexicon. If some of them are, compare your understanding of them with the definition that Word Spy gives.
5. Of all the original words created by J. K. Rowling in the Harry Potter books, the only words that have so far made their way into the online *Oxford English Dictionary* are the words *muggle*, *quidditch*, *Butterbeer*, and *hippogriff*. Look up the meanings of these words in the *OED* and speculate on why they have made it into the *OED* as opposed to other Harry Potter words like *house elf*, *bludger*, or *dementor*.

3 Lexeme Formation: The Familiar

CHAPTER OUTLINE

KEY TERMS

derivation
affixation
compounding
conversion
coinage
blending
backformation
corpus
formative
cran morph
extender
sound
symbolism
splinter

In this chapter you will learn about common ways of creating new lexemes.

- ◆ We will look at derivational affixation, considering the distinction between affixes and bases, and between free and bound bases.
- ◆ We will learn how to segment words into morphemes, how to formulate word formation rules, and how to determine the structure of words.
- ◆ We will consider what morphemes mean.
- ◆ Beyond affixation, in this chapter we will learn about processes of compounding, conversion, and other ways of creating new words.
- ◆ And you will learn to use corpora to search for your own morphological data.

3.1 Introduction

Take a look at the words below:

- unwipe (v.)
- converse (v.)
- googical (a.)
- fiberhood (n.)
- nympho-ology (n.)
- snorgle (v.)
- crankypants (n.)
- swapoutable (a.)

Have you ever heard these words before? Can you imagine what they mean?

Chances are that you haven't heard or read them before. Nevertheless, you probably didn't have much trouble figuring out at least roughly what their meanings might be. Assuming that you know the noun *conversation*, you can surely guess that *converse* must mean 'to have a conversation'. *Googical* must mean 'pertaining to Google', and in the phrase *Googical proportions*, metaphorically must mean 'huge'. The verb *snorgle* is harder, but you might guess that it's probably onomatopoeic.¹ And *crankypants*, *fiberhood*, and *swapoutable* are pretty transparent if you know what the individual morphemes mean.

The reason you can make educated guesses about these words is that they follow the rules of word formation in English. Once you know what the base – the central bit of the word – means, you can often figure out everything else. In this chapter, we're going to look at the most common ways of forming new lexemes, concentrating on English, but occasionally looking at other languages of the world.² You'll learn how to analyze words into their component parts, see how those parts are organized, and how the various parts contribute to their meanings.

3.2 Kinds of Morphemes

Most native speakers of English will recognize that words like *Googical*, *crankypants*, or *fiberhood* are made up of several meaningful pieces, and will be able to split them into those pieces:

- (1) un/wipe
 Google/ical
 cranky/ pants
 fiber/hood

As you learned in [Chapter 1](#), these pieces are called **morphemes**, the minimal meaningful units that are used to form words. Some of the

1. It's the noise that a small dachshund makes when clearing its throat.

2. We will look a lot more at languages other than English in [Chapter 5](#).

morphemes in (1) can stand alone as words: *wipe*, *Google*, *cranky*, *pants*, *fiber*. These are called **free** morphemes. The morphemes that cannot stand alone are called **bound** morphemes. In the examples above, the bound morphemes are *-ical*, *un-*, and *-hood*. Bound morphemes come in different varieties. Those in (1) are **prefixes** and **suffixes**; the former are bound morphemes that come before the base of the word, and the latter bound morphemes that come after the base. Together, prefixes and suffixes can be grouped together as **affixes**.³

New lexemes that are formed with prefixes and suffixes on a base are often referred to as **derived** words, and the process by which they are formed as **derivation**. The **base** is the semantic core of the word to which the prefixes and suffixes attach. For example, *wipe* is the base of *unwipe*, and *Google* is the base of *Googlical*. Frequently, the base is a free morpheme, as it is in these two cases. But stop a minute and consider the data in the next Challenge box.

Challenge

Divide the following words into morphemes:

- pathology
- osteopath
- dermatitis
- endoderm

Chances are that you recognize that there are two morphemes in each word. However, neither part is a free morpheme. Do we want to call these morphemes prefixes and suffixes? Would this seem odd to you?

If you said that it would be odd to consider the morphemes in our Challenge as prefixes and suffixes, you probably did so because this would imply that words like *pathology* and *osteopath* are made up of nothing but affixes!

Morphologists therefore make a distinction between affixes and **bound bases**. Bound bases are morphemes that cannot stand alone as words, but are not prefixes or suffixes. Sometimes, as is the case with the morphemes *path* or *derm*, they can occur either before or after another bound base: *path* precedes the base *ology*, but follows the base *osteo*; *derm* precedes another base in *dermatitis* but follows one in *endoderm*. This suggests that *path* and *derm* are not prefixes or suffixes: there is no such thing as an affix which sometimes precedes its base and sometimes follows it. But not all bound bases are as free in their placement as *path*; for example, *osteo* and *ology* seem to have more fixed positions, the former usually preceding another bound base, the latter following. Similarly, the base *-itis* always follows, and *endo-* always precedes another base. Why not call them respectively a prefix and a suffix, then?

3. We will see in Chapter 5 that there are other types of affixes as well.

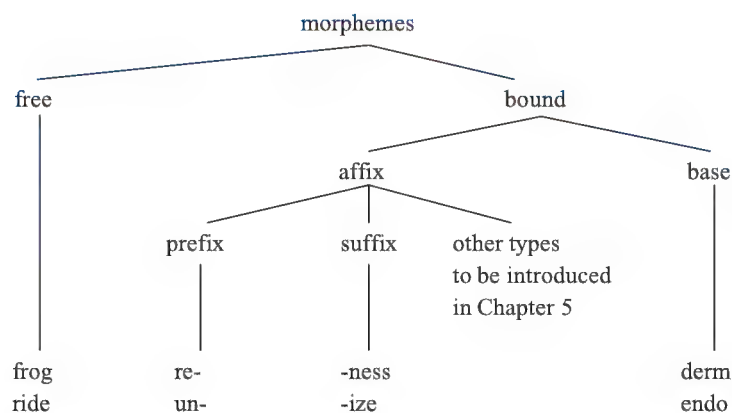


FIGURE 3.1
Types of morphemes

One reason is that all of these morphemes seem in an intuitive way to have far more substantial meanings than the average affix does. Whereas a prefix like *un-* (*unhappy*, *unwise*) simply means ‘not’ and a suffix *-ish* (*red-dish*, *warmish*) means ‘sort of’, *osteo* means ‘having to do with bones’, *-ology* means ‘the study of’, *path* means ‘sickness’, *derm* means ‘skin’, and *-itis* means ‘disease’. Semantically, bound bases can form the core of a word, just as free morphemes can. Figure 3.1 summarizes types of morphemes. We’ll look more carefully at the meanings of affixes in Section 3.3.

Another reason to believe that bound bases are different from prefixes and suffixes is that prefixes and suffixes tend to occur more freely than bound bases do. For example, any number of adjectives can be made negative by using the prefix *un-*, but there are far fewer words with the bound base *osteo*. This is perhaps not the best way of distinguishing between bound bases and affixes, though, as there are a few bound bases – *-ology* is one of them – that occur with great freedom, and there are some prefixes and suffixes that don’t occur all that often (e.g., the *-th* in *width* or *health*). So we’ll stick with the criterion of ‘semantic robustness’ for now. We’ll return in the next chapter to the question of how freely various morphemes are used in word formation.

With regard to bases, another distinction that’s sometimes useful in analyzing languages other than English is the distinction between **root** and **stem**. In languages with more inflection than English, there is often no such thing as a **free base**: all words need some sort of inflectional ending before they can be used. Or to put it differently, all bases are bound. Consider the data below from Latin:

- | | | | | | | | |
|-----|-------|---------|---------|----------|-----|---------------|-----------|
| (2) | Latin | 1st sg. | am + o | ‘I love’ | pl. | am + a + mus | ‘we love’ |
| | | | dic + o | ‘I say’ | | dic + i + mus | ‘we say’ |

In the singular, an ending signaling the first person (“I”) can sometimes attach to the smallest bound base meaning ‘love’ or ‘say’; this morpheme is the root. In the first person plural, and in most other persons and numbers, however, another morpheme must be added before the inflection goes on. This morpheme (an *a* for the verb ‘love’ and an *i* for the verb ‘say’) doesn’t mean anything, but still must be added before the

inflectional ending can be attached. The root plus this extra morpheme is the stem. Thought of another way, the stem is usually the base that is left when the inflectional endings are removed. We will look further at roots and stems in [Chapter 6](#), when we discuss inflection more fully.

3.3 Affixation

3.3.1 Word Formation Rules

Let's look more carefully at words derived by **affixation**. Prefixes and suffixes usually have special requirements for the sorts of bases they can attach to. Some of these requirements concern the phonology (sounds) of their bases, and others concern the semantics (meaning) of their bases – we will return to these shortly – but the most basic requirements are often the syntactic part of speech or **category** of their bases. For example, the suffix *-ness* attaches freely to adjectives, as the examples in (3a) show and sometimes to nouns (as in (3b)), but not to verbs (3c):

- (3) a. *-ness* on adjectives: redness, happiness, wholeness, commonness, niceness
- b. *-ness* on nouns: appleness, babeness, couch-potatiness⁴
- c. *-ness* on verbs: *runness, *wiggleness, *yawnness⁵

The prefix *un-* attaches to adjectives (where it means 'not') and to verbs (where it means 'reverse action'), but not to nouns:

- (4) a. *un-* on adjectives: unhappy, uncommon, unkind, unserious
- b. *un-* on verbs: untie, untwist, undress, unsnap
- c. *un-* on nouns: *unchair, *unidea, *ungiraffe

We might begin to build some of the rules that native speakers of English use for making words with *-ness* or *un-* by stating their categorial requirements:

- (5) Rule for *-ness* (first version): Attach *-ness* to an adjective or to a noun.
- Rule for *un-* (first version): Attach *un-* to an adjective or to a verb.

Of course, if we want to be as precise as possible about what native speakers know about forming words with these affixes, we should also indicate what category of word results from using these affixes, and what the resulting word means. So a more complete version of our *-ness* and *un-* rules might look like (6):

- (6) Rule for *-ness* (second version): *-ness* attaches to adjectives or nouns 'X' and produces nouns meaning 'the quality of X'.

Rule for *un-* (second version): *un-* attaches to adjectives meaning 'X' and produces adjectives meaning 'not X'; *un-* attaches to verbs meaning 'X' and produces verbs meaning 'reverse the action X'.

4. Some speakers will find the forms in (3b) odd, and will question their acceptability, but they are all attested in the Corpus of Contemporary American English and discussed in [Bauer, Lieber, and Plag \(2013\)](#).

5. Here and elsewhere, I use * before a word or sentence to indicate that it's unacceptable.

If we're really trying to model what native speakers of English know about these affixes, we might try to be even more precise. For example, *un-* does not attach to all adjectives or verbs, as you can discover by looking at the next Challenge box.

Challenge

Look at the following words and try to work out more details of the rule for *un-* in English. The (a) list contains some adjectives to which negative *un-* can be attached and others which seem impossible or at least somewhat odd. The (b) list contains some verbs to which reversible *un-* can attach and others which seem impossible. See if you can discern some patterns:

- (a) unhappy, *unsad, unlovely, *unugly, unintelligent, *unstupid
- (b) untie, unwind, unhinge, unknot, *undance, *unyawn, *unexplode, *unpush

What the (a) examples in the Challenge box seem to show is that the negative prefix *un-* in English prefers to attach to bases that do not themselves have negative connotations. This is not true all of the time – adjectives like *unselfish* or *unhostile* are attested in English – but it's at least a significant tendency. As for the (b) examples, they suggest that the *un-* that attaches to verbs prefers verbal bases that imply some sort of result, and moreover that the result is not permanent. Verbs like *dance*, *push*, and *yawn* denote actions that have no results, and although *explode* implies a result (i.e., something is blown up), it's a result that is permanent. In contrast, a verb like *tie* implies a result (something is in a bow or knot) which is temporary (you can take it apart).

We have just constructed what morphologists call a word **formation rule**, a rule which makes explicit all the categorial, semantic, and phonological information that native speakers know about the kind of base that an affix attaches to and about the kind of word it creates. We might now state the full word formation rules for negative *un-* as in (7):

- (7) Rule for negative *un-* (final version): *un-* attaches to adjectives, preferably those with neutral or positive connotations, and creates negative adjectives. It has no phonological restrictions.

Now let's look at two more affixes. In English we can form new verbs by using the suffixes *-ize*⁶ or *-ify*. Both of these suffixes attach to either nouns or adjectives, resulting in verbs:

- (8) *-ize* on adjectives: civilize, idealize, finalize, romanticize, tranquilize
- ize* on nouns: unionize, crystallize, hospitalize, caramelize, animalize

6. I use American English spelling here and elsewhere. This suffix would be spelled *-ise* in British English orthography.

-ify on adjectives: purify, glorify, uglify, moistify, diversify

-ify on nouns: mummify, speechify, classify, brutify, scarify, bourgeoisify

We might state the word formation rules for *-ize* and *-ify* as in (9):

- (9) Rule for *-ize* (first version): *-ize* attaches to adjectives or nouns that mean 'X' and produces verbs that mean 'make/put into X'.

Rule for *-ify* (first version): *-ify* attaches to adjectives or nouns that mean 'X' and produces verbs that mean 'make/put into X'.

But again, we can be a bit more precise about these rules. Although *-ize* and *-ify* have almost identical requirements for the category of base they attach to and produce words with roughly the same meaning, they have somewhat different requirements on the phonological form of the stem they attach to. As the examples in (8) show, *-ize* prefers words with two or more syllables where the final syllable doesn't bear primary stress (e.g., *tránquil*, *hóspital*). The suffix *-ify*, on the other hand, prefers monosyllabic bases (*pure*, *brute*, *scar*), although it also attaches to bases that end in a *-y* (*mummy*, *ugly*) or bases whose final syllables are stressed (*divérse*, *bourgéóis*). Since we want to be as precise as possible about our word formation rules for these suffixes, we will state their phonological restrictions along with their categorial needs:

- (10) Rule for *-ize* (final version): *-ize* attaches to adjectives or nouns of two or more syllables where the final syllable does not bear primary stress. For a base 'X' it produces verbs that mean 'make/put into X'.

I leave it to you to come up with the final version of the word formation rule for *-ify*.

3.3.2 Word Structure

When you divide up a complex word into its morphemes, as in (11), it's easy to get the impression that words are put together like the beads that make up a necklace – one after the other in a line:

- (11) unhappiness = un + happy + ness

But morphologists believe that words are more like onions than like necklaces: onions are made up of layers from innermost to outermost. Consider a word like *unhappiness*. We can break this down into its component morphemes *un* + *happy* + *ness*, but given what we learned above about the properties of the prefix *un-* and the suffix *-ness* we know something more about the way in which this word is constructed beyond just its constituent parts. We know that *un-* must first go on the base *happy*. *Happy* is an adjective, and *un-* attaches to adjectives but does not change their category. The suffix *-ness* attaches only to adjectives and makes them into nouns. So if *un-* attaches first to *happy* and *-ness* attaches next, the requirements of both affixes are met. But if we were

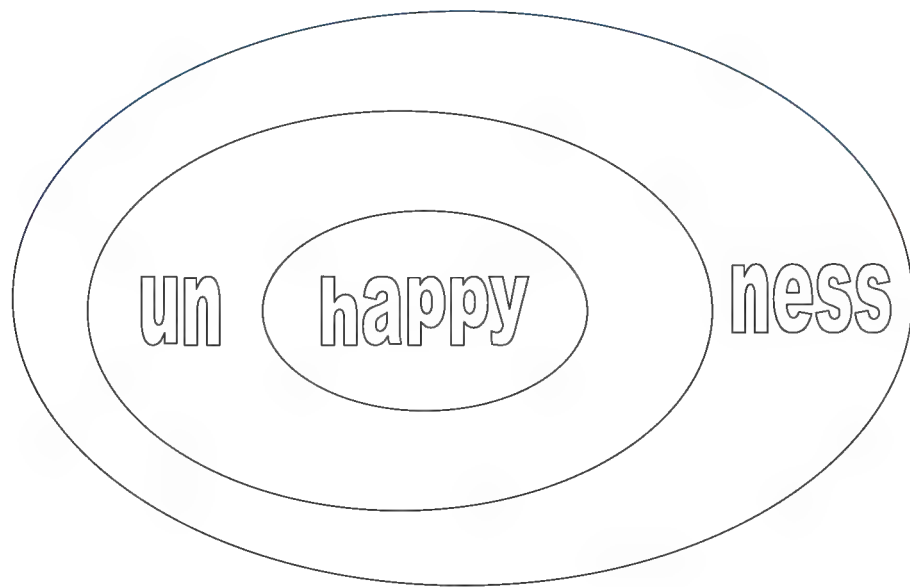
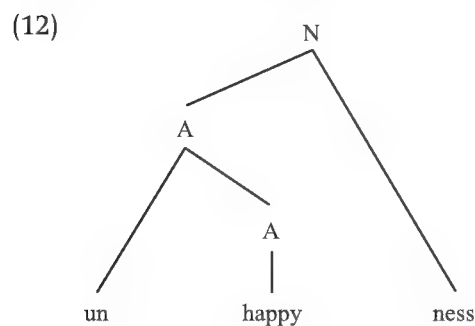


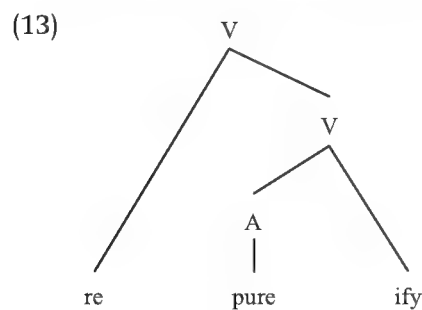
FIGURE 3.2
Words are like onions

to do it the other way around, *-ness* would have first created a noun, and then *un-* would be unable to attach. We could represent the order of attachment as if words really were onions, with the base in the innermost layer, and each affix in its own succeeding layer: see Figure 3.2.

But linguists, not generally being particularly artistic, prefer to show these relationships as ‘trees’ that look like this:



Similarly, we might represent the structure of a word like *repurify* as in (13):



In order to draw this structure, we must first know that the prefix *re-* attaches to verbs (e.g., *reheat*, *rewash*, or *redo*) but not to adjectives (**repure*, **rehappy*) or to nouns (**rechair*, **retruth*). Once we know this, we

can say that the adjective *pure* must first be made into a verb by suffixing *-ify*, and only then can *re-* attach to it.

Challenge

In English, the suffix *-ize* attaches to nouns or adjectives to form verbs. The suffix *-ation* attaches to verbs to form nouns. And the suffix *-al* attaches to nouns to form adjectives. Interestingly, these suffixes can be attached in a recursive fashion: *convene* → *convention* → *conventional* → *conventionalize* → *conventionalization*.

First draw a word tree for *conventionalization*. Then see if you can find other bases on which you can attach these suffixes recursively. What is the most complex word you can create from a single base that still makes sense to you? Are there any limits to the complexity of words derived in this way?

3.3.3 What Do Affixes Mean?

When we made the distinction between affixes and bound bases above, we did so on the basis of a rather vague notion of semantic robustness; bound bases in some sense had more semantic meat to them than affixes did. Let us now attempt to make that idea a bit more precise by looking at typical meanings of affixes.

In some cases, affixes seem to have not much meaning at all. Consider the suffixes in (14):

- (14) a. *-(a)tion* examination, taxation, realization, construction
 -ment agreement, placement, advancement, postponement
 -al refusal, arousal, disposal
- b. *-ity* purity, density, diversity, complexity
 -ness happiness, thickness, rudeness, sadness

Beyond turning verbs into nouns with meanings like ‘process of X-ing’ or ‘result of X-ing’, where X is the meaning of the verb, it’s not clear that the suffixes *-(a)tion*, *-ment*, and *-al* add much of any meaning at all. Similarly with *-ity* and *-ness*, these don’t carry much semantic weight of their own, aside from what comes with turning adjectives into nouns that mean something like ‘the abstract quality of X’, where X is the base adjective. Affixes like these are sometimes called **transpositional affixes**, meaning that their primary function is to change the category of their base without adding any extra meaning.

Contrast these, however, with affixes like those in (15):

- (15) a. *-ee* employee, recruit, deportee, inductee
 b. *-less* shoeless, treeless, rainless, supperless
 c. *re-* rehear, reread, rewash

These affixes seem to have more semantic meat on their bones, so to speak: *-ee* on a verb indicates a person who undergoes an action; *-less* means something like ‘without’; and *re-* means something like ‘again’.

Languages frequently have affixes (or other morphological processes, as we'll see in [Chapter 5](#)) that fall into common semantic categories. Among those categories are:

- **personal or participant affixes:** These are affixes that create 'people nouns' either from verbs or from nouns. Among the personal affixes in English are the suffix *-er* which forms **agent** nouns (the 'doer' of the action) like *writer* or *runner* and the suffix *-ee* which forms **patient** nouns (the person the action is done to). Also among this class of affixes are ones that create 'people nouns' other than the agent or the patient in an event, for example, inhabitants of a place (like *Manhattanite* or *Bostonian*).
- **locative affixes:** These are affixes that designate a place. For example, in English we can use the suffix *-ery* or *-age* to denote a place where something is done or gathered (like *eatery* or *orphanage*).
- **abstract affixes:** These are affixes that create abstract nouns that denote qualities (like *happiness* or *purity*) or statuses (*puppydom*, *advisors*hip, *daddyhood*) or even aspects of behavior (*buffoonery*).
- **negative and privative affixes:** Negative affixes add the meaning 'not' to their base; examples in English are the prefixes *un-*, *in-*, and *non-* (*unhappy*, *inattentive*, *non-functional*). Privative affixes mean something like 'without X'; in English, the suffix *-less* (*shoeless*, *hopeless*) is a privative suffix, and the prefix *de-* has a privative flavor as well (e.g., words like *debug* or *debone* mean something like 'cause to be without bugs/bones').
- **prepositional/relational affixes:** Prepositional/relational affixes often convey notions of space and/or time. Examples in English might be prefixes like *over-* and *out-* or *pre-* and *post-* (*overflow*, *overcoat*, *outrun*, *outhouse*, *preschool*, *preheat*, *postwar*, *postdate*).
- **quantitative affixes:** These are affixes that have something to do with amount. In English we have affixes like *-ful* (*handful*, *helpful*) and *multi-* (*multifaceted*). Another example might be the prefix *re-* that means 'repeated' action (*reread*), which we can consider quantitative if we conceive of a repeated action as being done more than once. Other quantitative affixes that we have in English denote collectives or aggregates of individuals (e.g., *acreage* or *knickknackery*).
- **evaluative affixes:** Evaluative affixes consist of **diminutives**, affixes that signal a smaller version of the base (e.g., in English *-let* as in *booklet* or *droplet*), and **augmentatives**, affixes that signal a bigger version of the base. The closest we come to augmentative affixes in English are prefixes like *mega-* (*megastore*, *megabite*). The Native American language Tuscarora (Iroquoian family) has an augmentative suffix *-ʔoʔy* that can be added to nouns to mean 'a big X'; for example, *takó:θ-ʔoʔy* means 'a big cat' ([Williams 1976](#): 233). Diminutives and augmentatives frequently bear other nuances of meaning. For example, diminutives often convey affection, or endearment, as we find in some words with *-y* or *-ie* in English (e.g., *sweetie*, *kitty*), although this is not necessarily the case (the Niger-Congo language Fula has a pejorative diminutive, for example).

Note that some semantically contentful affixes change syntactic category as well; for example, the suffixes *-er* and *-ee* change verbs to nouns, and the prefix *de-* changes nouns to verbs. But semantically contentful affixes need not change syntactic category. The suffixes *-hood* and *-dom*, for example, do not (*childhood*, *kingdom*), and *by* and large prefixes in English do not change syntactic category.

So far we have been looking at suffixes and prefixes whose meanings seem to be relatively clear. Things are not always so simple, though. Let's look more closely at the suffix *-er* in English, which we said above formed agent nouns. Consider the following words:

- (16) a. writer
skater
b. printer
freighter
c. loaner
fryer (i.e., a kind of chicken)
d. diner

All of these words seem to be formed with the same suffix. Look at each group of words and try to characterize what their meanings are. Does *-er* seem to have a consistent meaning?

It's rather hard to see what all of these have in common. The words in (16a) are indeed all agent nouns, but the (b) words are instruments; in other words, things that do an action. In American English the (c) words are things as well, but things that undergo the action rather than doing the action (like the patient *-ee* words discussed above): a *loaner* is something which is loaned (often a car, in the US), and a *fryer* is something (typically a chicken) which is meant to be fried. And the word *diner* in (d) denotes a location (a *diner* in the US is a specific sort of restaurant). Some morphologists would argue that there are four separate suffixes in English, all with the form *-er*. But others think that there's enough similarity among the meanings of *-er* words in all these cases to merit calling *-er* a single affix, but one with a cluster of related meanings. All of the forms derived with *-er* denote concrete nouns, either persons or things, related to their base verbs by participating in the action denoted by the verb, although sometimes in different ways. A cluster of related meanings like that exhibited by *-er* is called **affixal polysemy**.

Affixal polysemy is not unusual in the languages of the world. For example, it is not unusual for agents and instruments to be designated by the same suffix. This occurs in Dutch, as the examples in (17a) show (Booij and Lieber 2004), but also in Yoruba (Niger-Congo family), as the examples in (17b) show (Pulleyblank 1987: 978):

- (17) a. Dutch
spel-er 'player' (*spelen* 'play')
maa-er 'mower' (*maaien* 'mow')
b. Yoruba (Niger-Congo)
a-pàńà 'murderer' (pa 'kill', èmà 'people')
a-bẹ 'razor, penknife' (bẹ 'cut')

The Dutch suffix *-er* is in fact quite similar to the *-er* suffix in English in the range of meanings it can express. The Yoruba prefix *a-* also forms both agents and instruments.

3.3.4 To Divide or Not to Divide? A Foray into *Extenders, Formatives, Crans, and Other Messy Bits*

In [Chapter 1](#) we defined a morpheme as the smallest unit of language that has its own meaning. We have now looked at affixes and bases, both free and bound, and considered their meanings and how they combine into complex structures. For the most part, the examples we have looked at are simple and their analysis has been relatively clear. But often the closer we look at the morphology of a language, the more complex it becomes. There are many ways we can illustrate this with the morphology of English, but we will choose just a few points in this section that complicate our initial picture. I should point out in advance that my examples here will come from English, but we should expect to find similar twists and turns in any language whose morphology we examine in detail.

Consider the words in (18):

- (18) a. report, import, transport, deport, comport, export
- b. cranberry, huckleberry, raspberry
- c. Platonic, tobacconist, spasmodic, egotist
- d. sniffle, snort, snot, snout
- e. eggitarian, pizzatarian, pastatarian, fruitarian, flexitarian

Let's start with the ones in (18a). We have assumed so far that words can be broken down into morphemes, which are pieces that have meaning. The words in (18a) certainly look like they might be broken down because they have recurrent parts, but if they are morphemes, what do the pieces mean? In fact, English has dozens of words that are similar to what we might call the *-port* family. See how many cells of [Table 3.1](#) you can fill in.

Table 3.1	in-	ex-	con-	re-	trans-	de-
-port						
-mit						
-ceive						
-duce						
-cede						
-fer						
-scribe						
-gress						
-sist						

Challenge

Do you think that units like *-port*, *-mit*, *-ceive*, and the like should be considered morphemes? If so, what problems do they present for our definition of **morpheme**? If not, what should we do about the intuition that native speakers of English have that such words are complex?

One reason for our dilemma in analyzing these forms is that they are not native to English. They were borrowed from Latin (or from French, which in turn is descended from Latin), where they did have clear meanings: *-port* comes from the verb *portare* ‘to carry’, *-mit* from the verb *mittere* ‘to send’, *-scribe* from the verb *scribere* ‘to write’, and so on. But English speakers (unless they’ve studied Latin!) don’t know this. Morphologists are left with an unsatisfying sense that the words above somehow ought to be treated as complex, but are nevertheless reluctant to give up the strict definition of morpheme. One way of dealing with these pieces is to acknowledge that they seem to be independent and recombining in some way, but that they are not morphemes in the normal sense. Bauer, Lieber, and Plag (2013: 16) call elements like these **formatives**, which they define as “elements contributing to the construction of words whose semantic unity or function is obscure or dubious.”

The items in (18b) illustrate a different type of formative that are sometimes called **cran morphs**, from the first bit of the word *cranberry*. The second part of the word *cranberry* is clearly a free morpheme. But when we break it off, what’s left is a piece that doesn’t seem to occur in other words (except in recent years, words like *cranapple* that are part of product names), and doesn’t seem to mean anything independently. There are quite a few of these cran morphs in the names of other types of berries: *rasp-* in *raspberry*, *huckle-* in *huckleberry*. In cases such as these we are even more tempted than we were with *-port*, *-ceive*, and the like to divide words into morphemes, even though we know that one part of the word isn’t meaningful in the way morphemes usually are.

The examples in (18c) also display a puzzling characteristic. If we try to break these words down into their component morphemes, what we find is that each one consists of two obvious morphemes plus an extra sound or two:

- | | |
|---------------------|----------------------------------|
| (19) Plato + n + ic | (compare <i>icon</i> + ic) |
| tobacco + n + ist | (compare <i>accordian</i> + ist) |
| spasm + od + ic | (compare <i>Celt</i> + ic) |
| ego + t + ist | (compare <i>clarinet</i> + ist) |

The question is what the extra bit is. Is it part of the base of the word or part of the affix or part of neither? It seems pretty clear that it doesn’t mean anything. And why do we get an /n/ in *Platonic*, but /od/ in *spasmodic*, and nothing between the base and the suffix in *heroic*, or an /n/ in

tobacconist, but a /t/ in *egotist*? Morphologists don't have a clear answer to these questions – part of the fun of doing morphology is that we can argue about the possibilities, after all! – but we can at least give these puzzling bits a name. Following Bauer, Lieber, and Plag (2013), we will call them **extenders**.

Next, let's look at the examples in (18d). These exhibit what is called **sound symbolism**. All of the words begin with the consonant cluster /sn/ and seem to have something to do with the nose, but the sequence of sounds /sn/ doesn't mean anything by itself. Here morphologists are relatively agreed that sound symbolic words cannot be broken into parts. For one thing, the sequence /sn/ doesn't refer to 'nose' everywhere it occurs (consider words like *snail*, *snap*, or *snit*), and for another, if we were to segment /sn/ in the words in (18), what would be left over would neither have meaning by itself nor recur elsewhere in English.

Our final group of odd bits is illustrated in (18e). In these examples, the first part of each word is clearly a free morpheme, but the second part is not. Rather, it is what Bauer, Lieber, and Plag (2013) call a **splinter**, something which is split off from an original word, but which is not really (yet!) a true suffix. In the examples in (18e), the splinter is *tarian*, which is a bit broken off from the word *vegetarian*, and then used to create new words meaning 'one who eats X'. English has lots of splinters, among them *tastic*, as in *funktastic* or *fishtastic*, which is used to form mostly ironic words meaning 'excellent or great in reference to X', originally from *fantastic*, or *licious*, as in *bagelicious* or *bootielicious*, which is used to form words meaning 'appealing in reference to X', originally from the word *delicious*. The difference between a splinter and a true suffix is that speakers understand splinters in relation to the original word from which the ending splits off. If these bits survive and continue to give rise to new forms, though, they might someday be real suffixes!

3.3.5 How To: Dividing Words in English

Sometimes it's relatively simple to see how to divide up a complex word in English, but not always. In this section, I'll give you some ideas about how to think about the internal structure of complex words and about the issues that arise in trying to analyze them. A thought to keep in mind as we go along is that there isn't always a single correct way to divide up a word.

Let's consider first relatively simple words like those in (20):

- (20) a. nonunionization
b. nonparticipation

For (20a), we can feel quite certain that the base is the noun *union*, that there's a negative prefix *non-* and two suffixes, *-ize*, which takes nouns and makes verbs, and *-ation*, which takes verbs and makes nouns. For (20b), the prefix is still obviously the negative *non-*, but if we say that the suffix is *-ation*, as in (20a), and if we want to say that

-ation attaches to verbs to make nouns, we have a problem: peeling off -ation, we are left with something that looks like a bound base *particip* rather than a verb. English does, of course, have lots of bound bases (for example, *bapt* in *baptize*), but there's another way of thinking about this example. The alternative is to say that the base is the verb *participate*, and that the suffix, in this case is -tion. Analyzing the word this way, we don't have a bound base, but we do end up with two different forms of what looks to be the same suffix: -ation and -tion. Those two different forms are often called **allomorphs** – variants of morphemes. We will look a lot more at allomorphy in [Chapter 9](#). Problems of allomorphy like this crop up all the time in analyzing complex words in English.

Things can get murkier, though. Consider the word *monomaniacal*, which means something like 'pertaining to a mental illness focused on a single obsession or idea'. The first part of the word is clearly the prefix *mono-*, which adds the quantificational meaning 'one'. The rest of the word, though, leads us to several different possibilities:

- (21) a. mania + (a)c + al
 b. mania + (i)c + al
 c. mania + (i)cal

All three alternatives suggest that the base is the noun *mania*, which seems right. (21a) suggests that the next step turns *mania* into *maniac*. There are two questions about this possibility. First is -c or -ac a suffix in English? The OED recognizes it as a suffix meaning 'relating to or possessing a physical or mental condition' (as in, for example, *amnesiac* or *hemophiliac*) so this is a possibility for us. However, it might not be the best possibility: *maniac* doesn't usually mean 'a person suffering from a mania'. Rather, it typically has the special meaning 'raving with madness', which seems not to fit with the overall meaning of the word *monomaniacal*. If, however, we go with (21b) or (21c) we still have problems, though: do we split up the remainder of the word into two suffixes – -ic and -al – or do we say that -ical is a single suffix? The suffixes -ic and -al clearly are both suffixes that form adjectives, for example, *satiric* or *adenoidal*. But -ical also seems sometimes to function as a single unit – in a word like *whimsical*, there is no way of separating -ic from -al. Further, in some cases of words ending in the sequence -ical we have some evidence for splitting. For example, we have both *electric* and *electrical* as common words in English, sometimes used in subtly different ways. But we don't seem to be able to split *maniacal* up in this way. So here, I'd choose to go with the solution in (21c).⁷

Let's look at one final example. This one has to do not only with where to separate morphemes, but also with what to call bits of words. Let's take a familiar word like *psychology*. Clearly it can be divided into two

7. There's actually one more possibility! The segment -c- occasionally appears as an extender, as in *justification* or *application*. Could it be an extender here? What do you think?

bits, but there are problems both in segmenting the word and in figuring out what to call the bits. Consider the possibilities in (22):

- (22) a. psycho + logy
 b. psych + ology
 c. psycho + ology
 d. psych + o + logy

Historically, at least, the word *psychology* was formed from two neo-classical bound bases. The *-logy* or *-ology* part of the word still is a bound form, but these days it doesn't attach exclusively to other neo-classical bound bases; for example, in the word *garbageology*, it attaches to a free base, *garbage*. So is it still a bound base, or should we call it a suffix? As for the first part of *psychology*, the originally bound base has spun off in contemporary English into two free forms. One is *psych*, which is a shortened form of *psychology*, designating a kind of subject matter, class, or major. The other is *psycho*, a colloquial word for someone who is mentally ill or out of control. The latter meaning doesn't seem to be a part of the meaning of *psychology*, so we can probably rule out the segmentation in (22a) or (22c). That leaves (22b) and (22d) as possible segmentations, but it doesn't quite solve the naming question. Do we continue to call both bits bound bases, or do we call the first piece a free base and the latter a suffix? When does a bound base become a free form or an affix? These are questions to which I have no simple answer.

I will not try to resolve these issues here. My point is that there are questions that arise in trying to segment many words in English and in labeling the parts of words. Sometimes it's easy, but sometimes it's not. Many morphologists, pointing to the sorts of issues I've illustrated here, even deny that it makes sense to try to segment words into morphemes. We will return to all of these issues at various points in this book.

3.4 Compounding

So far we have concentrated on complex words formed by derivation, specifically by affixation. Derivation is not the only way of forming new words, of course. Many languages also form words by a process called **compounding**. **Compounds** are words that are composed of two (or more) bases, roots, or stems. In English we generally use free bases to compose compounds, as the examples in (23) show:

- (23) *English compounds*
 compounds of two nouns: windmill, dog bed, book store
 compounds of two adjectives: icy cold, blue-green, red hot
 compounds of an adjective and a noun: greenhouse, blackboard,
 hard hat
 compounds of a noun and an adjective: sky blue, cherry red,
 rock hard

3.4.1 When Do We Have a Compound?

How do we know that a sequence of words is a compound? Surprisingly, it's not that easy to come up with a single criterion that works in all cases.

In English, spelling is no help at all; unlike German, where compounds are always written as one word,⁸ in English there is no fixed way to spell a compound word. Some, like *greenhouse*, are written as one word, others like *dog bed*, as two words, and still others, like *producer-director* are written with a hyphen between the two bases.

A better criterion is stress; compounds in English are often stressed on their first or left-hand base, whereas phrases typically receive stress on the right. Compare, for example, a *gré'enhouse*, which is the place where plants are grown, to a *green hóuse*, that is, a house that's painted green. But it's not always the case that compounds are stressed on the left. For example, most people pronounce *apple píe* with stress on the second base, but *ápple cake* with stress on the left one. Yet we have the feeling that both are compounds; it seems illogical to consider one a compound and not the other.

There is, however, one test for identifying compounds that is fairly reliable: we can test for whether a sequence of bases is a compound by seeing if a modifying word can be inserted between the two bases and still have the sequence make sense. If a modifying word cannot sensibly be inserted, the sequence of two words is a compound. This test confirms that both *apple pie* and *apple cake* are compounds, in spite of their differing stress. In neither case can we insert a modifier like *delicious* between the two stems; **apple delicious pie* and **apple delicious cake* are equally peculiar!

3.4.2 Compound Structure

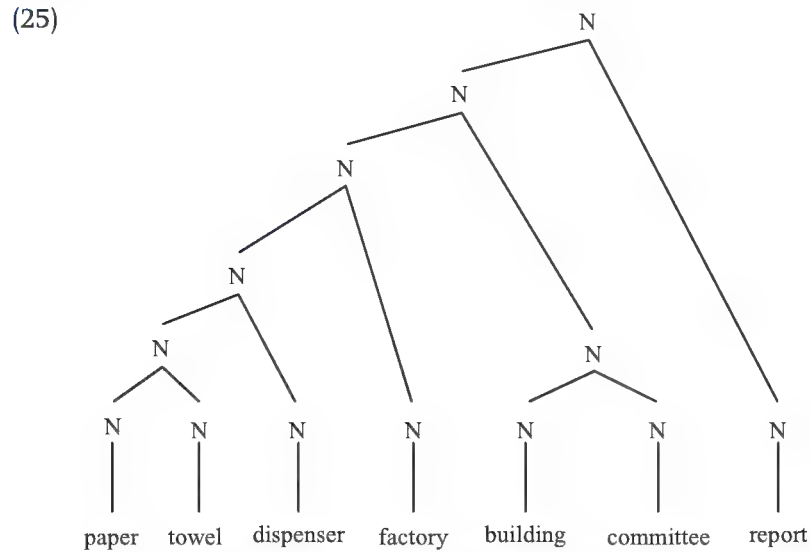
We can look at compounds as having internal structure in precisely the same way that derived words do, and we can represent that structure in the form of word trees. The compounds *windmill* and *hard hat* would have the structures in (24):



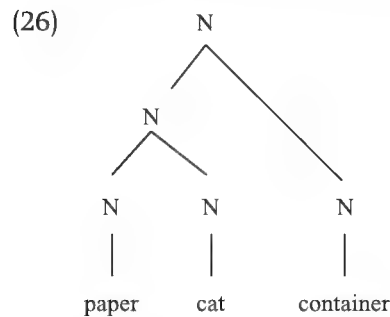
Compounds, of course, need not be limited to two bases. Compounding is what is called a recursive process, in the sense that a compound of two bases can be compounded with another base, and this compounded with still another base, so that we can eventually obtain very complex compounds like *paper towel dispenser factory building committee report*. As with derived words, it is possible to show the internal

8. English-speaking students of German often delight in compounds like *Donaudampfschiffahrtsgesellschaftskapitänsmütze* (Neef 2009), which means 'Cap of the captain of the Danube steam ship company'.

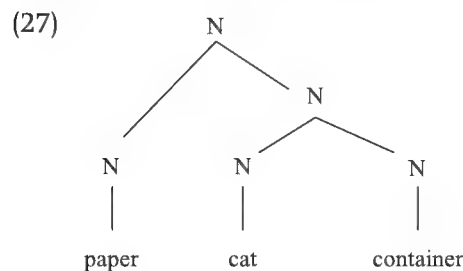
structure of complex compounds using word trees. Assuming that this compound is meant to denote a report from the building committee for the paper towel dispenser factory, we might give it the structure in (25):



Some compounds can be ambiguous, and therefore can be represented by more than one structure. For example, the compound *paper cat container*, might have this structure:



The way we've drawn this tree, the compound *paper cat* has been compounded with the noun *container* to make a complex compound. The compound as a whole then must mean 'a container for paper cats'. But if the compound *paper cat container* were intended to mean 'a paper container for cats', the structure of the tree would be that in (27), where *cat container* is first compounded, and then *paper* added in:⁹



9. These two interpretations are sometimes distinguished in spoken speech by placement of stress.

Often, the more complex the compound is, the greater the possibility of multiple interpretations, and therefore multiple structures.

Challenge

The compound *paper towel dispenser factory building committee report* could in fact have more than one meaning. See how many different meanings you can come up with, and draw a tree that corresponds to each of those meanings.

Languages other than English frequently construct compounds on free bases just as English does, although we can see in the French and Vietnamese examples in (28) that the order of elements in the compound is sometimes different from that in English, a fact we will return to in the [next section](#):

- (28) a. *French*:
 timbre poste ‘stamp-post = postage stamp’
 chêne liège ‘oak cork = cork oak’
- b. *Dutch*:
 boekhandel ‘book shop’
 zakgeld ‘pocket money’
- c. *Vietnamese*:
 nháthuong ‘establishment be-wounded = hospital’
 ngườioi ‘person be-located = servant’

As we saw above, English has bound bases as well as free bases, and when we put two of them together, as in the examples in (29), we might call these forms compounds as well. Some linguists call them **neo-classical compounds**, as the bound bases usually derive from Greek and Latin:

- (29) *English compounds on bound bases*: psychopath, pathology, endoderm, dermatitis

In languages like Latin where, as we saw, word formation often operates on roots or stems, rather than on free forms, all compounds are formed from bound bases. Specifically, the first parts of the compounds in (30) are formed from the roots of the nouns *ala* ‘wing’ and *capra* ‘goat’ (respectively *al-* and *capr-*) plus a vowel *-i-* linking the two parts of the compound together:

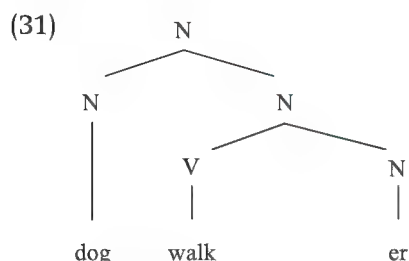
- (30) *Latin compounds*:
 ali-pes ‘wing-footed’
 capri-ficus ‘goat fig = wild fig’

The *-i-* that occurs between the two roots has no meaning, and is not the vowel that usually precedes the inflections (for these two nouns, that vowel would be *-a*). It is there solely to link the parts of the compound together,

and might be seen as another example of the bits called ‘extenders’ that we discussed in Section 3.3.4. Other morphologists call this bit a **linking element** or **interfix**.

3.4.3 Types of Compounds

In English and other languages there may be a number of different ways of classifying compounds. Traditionally, English compounds have been classified as either **root** (also known as **primary**) **compounds** or **synthetic** (also known as **deverbal**) **compounds**. Synthetic compounds are composed of two lexemes, where the second element is derived from a verb, and the first is interpreted as an argument of that verb.¹⁰ *Dog walker*, *hand washing*, and *home made* are all synthetic compounds.



Root compounds, in contrast, are made up of two lexemes,¹¹ which may be nouns, adjectives, or verbs; the second lexeme is typically not derived from a verb. The interpretation of the semantic relationship between the head and the non-head in root compounds is quite free as long as it's not the relationship between a verb and its argument. Compounds like *windmill*, *ice cold*, *hard hat*, and *red hot* are root compounds.

Although this is a traditional way of classifying English compounds, some morphologists like to use more fine-grained systems of classifications that focus on several distinct aspects of compound structure, and this is what we will do here. One advantage of doing this is that it makes it easier for us to compare compounds in English with compounds in other languages. In this section, we will discuss three different ways of classifying compounds: first, according to the part of speech or category of the words that make them up; second, according to the semantic relationships that they express; and third, according to the placement of what's called the **head** of the compound, a term that we will define shortly.

3.4.3.1 Classifying Compounds According to Category

One simple way of looking at compounds is to see what category or part of speech the parts of the compound belong to, and what part of speech the compound as a whole exhibits. In English, as we saw in (23), we can

10. We will go into arguments in more depth in Chapter 8. For now, it's enough to know that the arguments of the verb are its subject and its complements (direct object, indirect object, and so on).

11. Or sometimes more, as we saw in the last section.

form compounds from any combination of nouns and adjectives, but it's harder to form compounds using verbs or prepositions:

(32) Compound elements	Category of compound	Examples
N + N	N	dog bed, file cabinet, apple pie
N + A	A	sky blue, stone cold, bone dry
A + A	A	blue-green, icy cold
A + N	N	blackbird, greenhouse, fast food
N + V	V	hand wash, brainwash, babysit
V + N	N	pick pocket, sell sword, think tank
V + V	V	stir-fry, slam dunk, blow dry
A + V	V	hot glue, slow dance
V + A	A	go-slow
N + P	Adv?	year-in (Bauer et al. 2013: 452) ¹²
P + N	N	afterbirth, backseat
V + P	N	breakdown
P + V	V	downgrade
A + P	A	tuned-in
P + A	A	inbuilt
P + P	P	into

In the cases with prepositions and verbs, it is difficult to find plausible examples, and some morphologists would question whether we really can form compounds composed of a noun plus a preposition or a verb plus an adjective in English, for example.

3.4.3.2 Classifying Compounds According to Semantic Relationships

We can also classify compounds more closely according to the semantic and grammatical relationships holding between the elements that make them up. One useful classification is that proposed by [Bisetto and Scalise \(2005\)](#), which recognizes three types of relation. The first type is what might be called an **attributive compound**. In an attributive compound the non-head acts as a modifier of the head. So *snail mail* is (metaphorically) a kind of mail that moves like a snail, and a *windmill* is a kind of mill that is activated by wind. With attributive compounds the first element might express just about any relationship with the head. For example, a *schoolbook* is a book used at school, but a *yearbook* is a record of school activities over a year. And a *notebook* is a book in which one writes notes. With a new compound (one I've just made up) like *mud wheel*, we are free to come up with any reasonable semantic relationship between the two bases, as long as the first modifies the second in some way: a wheel used in the mud, a wheel made out of mud, a wheel covered in

12. [Bauer, Lieber, and Plag \(2013\)](#) give this as an example of an N + P compound, but don't say what the resulting category is. In the case of most compounds, it's clearly that of the right-hand category, but with this combination of categories it's not at all clear!

mud, and so on. Some interpretations are more plausible than others, of course, but none of these is ruled out.

In **coordinative compounds**, the first element of the compound does not modify the second; instead, the two have equal weight. In English, compounds of this sort can designate something which shares the denotations of both base elements equally, or is a mixture of the two base elements, or expresses a relationship between the two elements:

- (33) *Coordinative compounds*: producer-director, prince consort, blue-green, doctor-patient

A *producer-director* is equally a producer and a director, a *prince consort* at the same time a prince and a consort. In the case of *blue-green* the compound denotes a mixture of the two colors. Finally, there are also coordinative compounds that denote a relation between the two bases (like *doctor-patient* in *doctor-patient confidentiality*). We will return to these below.

We find a third kind of semantic/grammatical relationship in **subordinative compounds**. In subordinative compounds one element is interpreted as the argument of the other, usually as its object. Typically this happens when one element of the compound either is a verb or is derived from a verb, so the synthetic compounds we looked at at the very beginning of this section are subordinative compounds in English. Some more examples are given in (34):

- | | | |
|------|--------------------|--|
| (34) | with- <i>er</i> | truck driver, mystery writer, lion tamer |
| | with- <i>ing</i> | truck driving, food shopping, hand holding |
| | with- <i>ation</i> | meal preparation, home invasion |
| | with- <i>ment</i> | cost containment |

It is easy to see that subordinative compounds are interpreted in a very specific way: that is, the first element of the compound is most often interpreted as the object of the verb that forms the base of the deverbal noun: for example, a *truck driver* is someone who *drives trucks*, *food preparation* involves *preparing food*, and so on.

Synthetic compounds are not the only subordinate compounds, however. A second type of subordinate compound is poorly represented in English, but occurs with great frequency in Romance languages like Spanish, French, and Italian:

- | | | |
|------|----------------|---|
| (35) | <i>English</i> | pickpocket |
| | <i>Italian</i> | lava piatti (lit. 'wash dishes' = 'dishwasher') |
| | <i>Spanish</i> | saca corcho (lit. 'pull cork' = 'corkscrew') |

In these compounds the first element is a verb, and the second bears an argumental relationship to the first element, again typically the object relationship. We will return to these shortly.

3.4.3.3 Classifying Compounds According to Headedness

A third way of classifying compounds concerns whether or not they have what is called a head, and if they do have a head, according to where the

head is located. We start with a definition: the head of the compound is the element that serves to determine both the part of speech and the semantic kind denoted by the compound as a whole. For example, in English the base that determines the part of speech of compounds such as *greenhouse* or *sky blue* is always the second one; the compound *greenhouse* is a noun, as *house* is, and *sky blue* is an adjective as *blue* is. Similarly, the second base determines the semantic category of the compound – in the former case a type of building, and in the latter a color. English compounds are therefore said to be **right-headed**. In other languages, however, for example French and Vietnamese, the head of the compound can be the first or leftmost base. For example a *timbre poste* (36a) is a kind of stamp, and a *người ở* (36b) is a kind of person. French and Vietnamese can therefore be said to have **left-headed** compounds:

- (36) a. *French*:
 timbre poste ‘stamp-post = postage stamp’
 chêne liège ‘oak cork = cork oak’
 b. *Vietnamese*:
 nhà thương ‘establishment be-wounded = hospital’
 người ở ‘person be-located = servant’

Using the notion of head, we can further divide compounds into **endocentric** or **exocentric** varieties. In endocentric compounds, the referent of the compound is always the same as the referent of its head. So a *windmill* is a kind of mill, and a *truck driver* is a kind of driver. Endocentric compounds are illustrated in (37):

- (37) *Endocentric compounds*
 Attributive: windmill, greenhouse, sky blue, icy cold
 Subordinative: truck driver, meal preparation

The French, and Vietnamese compounds in (36) are endocentric, as well, although as we pointed out above, the head occurs on the left in these compounds.

Coordinative compounds like *producer-director* or *blue-green* present a bit of a complication for us here. On the one hand, it’s certainly true that a *producer-director* is a kind of director, but he or she is also a kind of producer, and *blue-green* is sort of a mixture between blue and green. We might reasonably say then that these compounds have two semantic heads, in which case we might still be justified in considering them endocentric.

Compounds may be termed exocentric when the denotation of the compound as a whole is not the denotation of the head. For example, the English attributive compounds in (38) all refer to types of people – specifically stupid or disagreeable people – rather than types of heads, brains, or clowns, respectively. So *air heads* are people with nothing but air in their heads, and so on.

- (38) *Exocentric compounds*
 Attributive: air head, meat head, bird brain, ass clown

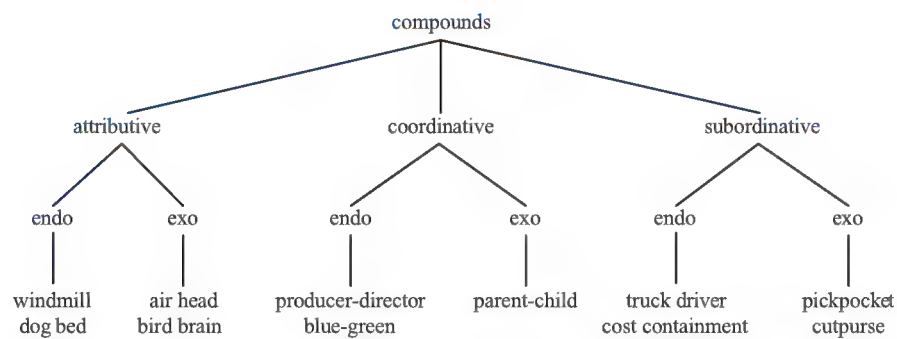


FIGURE 3.3:
Types of compounds

Coordinative: parent–child, doctor–patient

Subordinative: pickpocket, cutpurse, lava piatti (Italian, lit. ‘wash dishes’)

In coordinative compounds like *parent–child* or *doctor–patient* the compound as a whole denotes a relationship between its elements; when we speak of a *parent–child conversation* or a *doctor–patient conference*, the compound does not have a single referent, and some would argue that it’s not even a noun, even though it is made up of two nouns. So the coordinative examples in (38) would be exocentric as well. Finally, we saw examples of exocentric subordinative compounds from English, Spanish, and Italian in (35). English has only a few examples: *a pickpocket* is not a type of pocket, but a sort of person (who picks pockets). Romance languages have many compounds of this type, however.

The different types of compounds are summarized in Figure 3.3.

3.5 Conversion

Although we often form new lexemes by affixation or compounding, in English it is also possible to form new lexemes merely by shifting the category or part of speech of an already existing lexeme without adding an affix. This means of word formation is often referred to as **conversion** or **functional shift**. In English, we often create new verbs from nouns, as the examples in (39a) show, but we also do the reverse (39b), and sometimes we can even create new verbs from adjectives (39c):

- (39) a. table to table
 bread to bread
 fish to fish
- b. to throw a throw
 to kick a kick
 to fix a (quick) fix
- c. cool to cool
 yellow to yellow

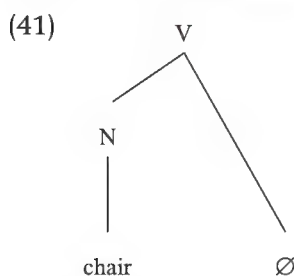
When we create new verbs from nouns, the resulting verbs may have a wide range of meanings. For example, *to bread* is ‘to put bread (crumbs) on something’, but *to fish* is ‘to take fish from a body of water’. And *to clown* is ‘to act like a clown’ rather than to put a clown somewhere or take a clown from somewhere! Going in the opposite direction, the meaning of the new word is somewhat more predictable; that is, when we turn a verb into a noun, the result often means something like ‘an instance of X-ing’, where X is the denotation of the verb. So for example, *a throw* is ‘an instance of throwing’.

English is, of course, not the only language with conversion. Noun to verb conversion occurs frequently in German and Dutch as well, as the examples in (40a–b) show, and verb to noun conversion is said to occur in French, as the examples in (40c) show:

- (40) a. *German*¹³
- | | | | |
|---------|----------------|------------|-------------|
| Antwort | ‘answer’ | antwort-en | ‘to answer’ |
| Strick | ‘cord, string’ | strick-en | ‘to knit’ |
- b. *Dutch*
- | | | | |
|--------|-----------|-----------|--------------|
| fiets | ‘bicycle’ | fiets-en | ‘to bicycle’ |
| hamer | ‘hammer’ | hamer-en | ‘to hammer’ |
| winkel | ‘shop’ | winkel-en | ‘to shop’ |
- c. *French*
- | | | | |
|----------|------------|--------|---------|
| gard-er | ‘to guard’ | garde | ‘guard’ |
| visit-er | ‘to visit’ | visite | ‘visit’ |

There may appear to be a suffix added in the derivation of the verbs in the examples in (40a–b) and one deleted in (40c). But the *-en* suffix in German and Dutch and the *-er* suffix in French do not derive the verbs per se – they are inflectional morphemes that signal the infinitive form of the verb. If we assume that conversion involves only the base or root, these examples count as conversion.

Morphologists have been divided on how to analyze conversion. Some argue that conversion is just like affixation, except that the affix is phonologically null – that is, it is unpronounced. When analyzed this way, conversion is called **zero affixation**. It might be represented structurally as in (41).



Other morphologists argue that conversion is different from affixation, and treat it simply as change of category with no accompanying change

13. Nouns are capitalized in German orthography.

of form, as we have done here. With this analysis, converted verbs like *to chair* would not have any internal structure, but would simply be regarded as having been relisted or recategorized in our mental lexicons. We will not decide between these analyses here.

Challenge

Is it possible in English for already compounded or affixed words to undergo conversion? Try to think of examples of words with prefixes or suffixes or compound words that can function as more than one part of speech (e.g., as both nouns and verbs).

3.6 Marvelous Intricacies: How Affixation, Compounding, and Conversion Interact

You might think that in creating new words we might be content to take a base or two and make a compound or add a prefix or suffix and leave it at that. But native speakers of English (and of course other languages) are remarkably exuberant in the way that they combine different types of word formation strategies to create wonderfully complex words. Consider, for example, the words in (42).

- (42)
- a. *multiple suffixes*: weaklingdom, ducklingish, institutionalize
 - b. *multiple prefixes*: hypermultilateralize, super-mega-omni-prayer, micronanosecond
 - c. *multiple prefixes and suffixes*: reinstitutionalization, counterreformational
 - d. *suffixes on compound words*: bestsellerdom, basketballer, dotcomist, girlfriendless
 - e. *prefixes on compound words*: ex-ballplayer, mega-earthquake, unfootnoted
 - f. *compound words that have undergone conversion*: to blacklist, to breakfast, a stir-fry,

Such words show us that we can form new words with several prefixes or suffixes or a mixture of the two and that both simple bases and compound bases may take prefixes and suffixes. And we can even take compounds of one category and change them to another using conversion, as the examples in (42f) suggest.

3.7 Minor Processes

Affixation, compounding, and conversion are the most common ways of forming new words, at least in English (we will see in [Chapter 5](#) that there are other means of word formation that languages other than

English use). In addition, there are a number of less common ways in which new lexemes may be formed. We provide a survey of them here, without going into great depth on any one of them.

3.7.1 Coinage

It is of course possible to make up entirely new words from whole cloth, a process called **coinage**. However, it turns out that we coin completely new words relatively rarely, choosing instead to recycle bases and affixes into new combinations. New products are sometimes given coined names like *Kodak*, *Xerox*, or *Kleenex*, and these in turn sometimes come to be used as common nouns: *kodak* was at one time used for cameras in general, and *xerox* and *kleenex* are still used respectively for copiers and facial tissue by some American English speakers. But it's relatively rare to coin new words. In hundreds of new words archived on the Word Spy website (www.wordspy.com), I was able to find only the following four apparent coinages:

- (43) blivet 'an intractable problem'
 mung 'to mess up, to change something so that it no longer works'
 grok 'to understand in a deep and exhaustive manner'
 (from Robert Heinlein's *Stranger in a Strange Land*)
 mongo 'objects retrieved from the garbage'

The verb *snorgle*, among the words with which we began this chapter, is probably another example. Why are there so few coinages? Perhaps because the words themselves give no clue to their meaning. Context often clarifies what a word is intended to mean, but without a context to suggest meaning, the words themselves are semantically opaque. It is no wonder that many of the pure coinages that creep into English come from original product names: the association of the coined word with the product makes its meaning clear, and occasionally the word will then be generalized to any instance of that product, even if manufactured by a different company.

3.7.2 Backformation

Generally, when we derive words we attach affixes to bases; in other words, the base exists before the word derived by affixation. For example, we start with the verb *write* and form the agent noun *writer*. Sometimes, however, there are words that historically existed as monomorphemic bases, but which ended in a sequence of sounds identical to or reminiscent of that of certain affixes. When native speakers come to perceive these words as being complex rather than simple, they create what is called a **backformation**. For example, historically the word *burglar* was monomorphemic. But because its last syllable was phonologically identical to the agentive *-er* suffix, some English speakers have understood it to be based on a verb *to burgle*. Arguably for those speakers, then, *burglar* is no longer a simple word. Similarly, the verb *surveil* has

been created from *surveillance* and the verb *liaise* from *liaison*. At least at first, some native speakers will find the backformations odd-sounding or objectionable. Some years ago I heard the then governor of Iowa, Tom Vilsack, use the verb *incent* on National Public Radio; in context, it clearly was a backformation from the noun *incentive*, and it sounded quite odd at the time. But with time, that feeling of oddness will disappear. Indeed speakers are sometimes surprised to learn that the verb did not exist before the corresponding noun, so ordinary-sounding has the verb come to be. Such is the case for *peddle* and *edit*, both of which are historically backformations from *peddler* and *editor* respectively.

3.7.3 Blending

Blending is a process of word formation in which parts of lexemes that are not themselves morphemes are combined to form a new lexeme. Familiar examples of **blends** (sometimes also called **portmanteau words**) are words like *brunch*, a combination of *breakfast* and *lunch*, or *smog*, a combination of *smoke* and *fog*. While not one of the major ways of forming new words, blending is used quite a bit in English in advertizing, product-naming, and playful language. The Word Spy website lists these blends:

(44) skitch	‘to propel oneself while on a skateboard or in-line skates by hanging onto a moving vehicle’ (combination of <i>skate</i> and <i>hitch</i>)
spime	‘a theoretical object that can be tracked precisely in space and time over the lifetime of the object’ (combination of <i>space</i> and <i>time</i>)
splog	‘a fake blog’ (combination of <i>spam</i> and <i>blog</i>)
vortal	‘a vertical portal’
bagonize	‘to wait anxiously for one’s bag to appear on the carousel at the airport’ (combination of <i>bag</i> and <i>agonize</i>)
Chrismukkah	‘a holiday celebration that combines elements of Christmas and Hanukkah’

Indeed, the sheer number of words of this sort that can be found in the Word Spy archives suggests the vitality of this process. Note that while most of the time the parts that are fused together to form blends are not themselves morphemes, sometimes a whole base or affix will be used; for example, Word Spy also lists the word *celeblog* (‘a blog written by a celebrity’) which is made up of the chunk *celeb* from *celebrity* and the word *blog*; the latter part has become a free morpheme in English in the last few years.

3.7.4 Acronyms and Initialisms

When the first letters of words that make up a name or a phrase are used to create a new word, the results are called **acronyms** or **initialisms**. In acronyms, the new word is pronounced as a word, rather than as a series

of letters. For example, Acquired Immune Deficiency Syndrome gives us *AIDS*, pronounced [eidz]. And self-contained underwater breathing apparatus gives us *scuba*. Note in the case of *scuba*, the acronym has become so familiar to English speakers that many do not know that it's an acronym! A favorite acronym of mine is the name for a now-defunct supermarket called the Durham Market Place, which was universally referred to as the DUMP.

Initialisms are similar to acronyms in that they are composed from the first letters of a phrase, but unlike acronyms, they are pronounced as a series of letters. So most people in the US refer to the Federal Bureau of Investigation as the *FBI* pronounced [ɛf bi ɑɪ]. Other initialisms are *PTA* for Parent Teacher Association, *PR* for either 'public relations' or 'personal record', and *NCAA* for National College Athletic Association.

3.7.5 Clipping

Clipping is a means of creating new words by shortening already existing words. For example, we have *info* created from *information*, *blog* created from *web log*, or *fridge* from *refrigerator*. Universities are fertile grounds for the creation of clippings: students study *psych*, *anthro*, *soc*, and even *ling* with one *prof* or another, and if they're taking a science class, may spend long hours in the *lab*, which might or might not involve running some *stats*. Although clippings are often used in a colloquial rather than a formal register, some have attained more neutral status. The word *lab*, for example, is probably used far more frequently in the US than its longer version *laboratory*. The word *mob* is a seventeenth-century clipping from the Latin term *mobile vulgus* 'the fickle common people'; the Latin phrase has long been forgotten, but the clipping persists as the normal word for an unruly throng of people.

3.8 How To: Finding Data for Yourself

One of the most exciting parts of learning about morphology is that we have all kinds of data at our fingertips, and with only a bit of training we can begin to find these data and come up with hypotheses of our own. In this section I will introduce you to a technique you can use to explore the morphology of English in more depth and begin to make discoveries of your own.

Morphologists have always used their own intuitions to find data – for example, trying to think of words that contain a particular prefix or suffix, or trying to form particular kinds of compounds. But sometimes it's hard to think of pertinent examples and sometimes the examples we come up with might sound funny or not quite right. We can of course go to dictionaries to check whether those words occur, or to try to find other words with the characteristics we're searching for, but we already know that the complex words we find in dictionaries are not necessarily representative of the complex words we can form.

Challenge

Take five minutes and see how many words you can come up with off the top of your head with the suffix *-ity*. Then take another five minutes to page through a dictionary (you'll need an old-fashioned paper dictionary for this challenge!) and see how many more *-ity* words you can come up with. How many words do you have?

paucity
 raucity
 caducity
 rumti-iddity
 rump-iddity
 quiddity
 oddity
 heredity
 rabidity
 morbidity
 turbidity
 acidity
 subacidity
 placidity
 nonacidity
 hypoacidity
 hyperacidity
 flaccidity
 rancidity

Chances are that no matter how many words you came up with, you might have a nagging feeling that you don't have enough words to really see how the suffix *-ity* works. But how to get more data?

Back in the dark ages before the internet, one way to look for words was to look in a backwards word list like [Lehnert \(1971\)](#). Backwards word lists were books that contained words alphabetized starting with the last letter, rather than the first. Theoretically in such lists all the words with *-ity* could be found together. A few of the *-ity* words to be found in [Lehnert \(1971: 584\)](#) are shown in the sidebar. You'll notice that using a backwards word list is not a perfect tool: such word lists simply alphabetize words from the end to the beginning, so any word ending in the letters *ity* will occur in the list, not just words that really have the suffix *-ity*. In the list I give here, in addition to real *-ity* words like *oddity* and *rancidity*, we find 'junk' like *rumti-iddity* (spelled two different ways!); a bit earlier in the list we would have found the word *city* which of course also ends in the sequence of letters *ity* but certainly doesn't contain a suffix.

At the time, using a backwards word list was an improvement over picking one's own brains for examples, but these days we have much better resources available to us in the form of vast databases of spoken and written language called **corpora** (singular **corpus**) that can be searched in quite sophisticated ways. Using corpora, you can not only find abundant words with any affix – prefix or suffix – but you can also see those words as they're used in context. Since some of the words that you encounter will not be recorded in dictionaries, you get a chance to see how speakers create new words and what those words mean in context.

For English, several corpora are freely available on the internet, and with just a bit of training you can use them to find all kinds of nifty data on contemporary English word formation. The best current source for these is the English-corpora.org website (www.english-corpora.org), which makes available several corpora, among them those in (45):

(45) English corpora available from English-corpora.org

- News on the Web (NOW) has over 100 billion words from a variety of news sources (2010–present, updated daily).
- Global Web-Based English (GloWbE) has over 1.9 billion words from 20 English-speaking countries (2012–2013).

- Corpus of Contemporary American English (COCA) has 1 billion words of American English from spoken language, newspapers, academic journals, magazines, and fiction (1990–2019).
- British National Corpus (BNC) has 100 million words of British English taken from spoken and written sources (1980–1993).
- Corpus of Historical American English (COHA) has 400 million words of American English taken from written sources (1810–2000s).

Each of these corpora has its own merits, depending on what you're looking for. COCA and the BNC are useful if you're specifically interested in American or British English. GloWbE is a great source if you want to compare English in different parts of the world. NOW is the largest and most up-to-date corpus. And COHA can give you some historical depth, if you're interested in looking back at how American English has changed.

Let's now look at how to use these corpora to find examples of prefixed or suffixed words. We'll use COCA for this demonstration, but the user interface is the same for all corpora on the English-corpora.org website, so once you've learned to use COCA, you'll be able to use all of them. Once you've found the English-corpora.org website, click on COCA and this will take you to a user interface that looks the one in [Figure 3.4](#).

To find words with a particular prefix or suffix, you'll need to do what's called a wildcard search. In a wildcard search, you use the asterisk symbol (*) to mean 'anything at all'. For example, if you want to search for words with the suffix *-esque*, you would enter **esque* in the "Find matching strings" box; **esque* means 'anything at all ending in the sequence of letters *esque*'. For a prefix like *mega-* you would enter *mega** to mean anything at all beginning with the sequence of letters *mega*, and so on. If you hit the search button now using the default settings, the corpus will give you the 100 most frequently occurring examples ending in that sequence of letters. This may give you a lot to work with, but for reasons that we will learn in the [next chapter](#), morphologists are often interested in the examples that occur the *least* frequently rather than those that occur *most* frequently. So it's good to learn how to fine-tune your search.

To find more examples and less frequently occurring ones, you will have to change the default settings. First click "Options," which is under



FIGURE 3.4
COCA website interface

the search box and change the setting that reads #HITS. You can set this for a much higher number than the default 100; I usually set this at 6,000, so unless there are a really huge number of ‘hits’, you should get the majority of words with the prefix or suffix you’re looking at. Your results will be sorted by frequency, which means that the words that occur most often will be at the top of your list of results. If you’re looking for infrequent forms, as morphologists often are, you’ll want to see the low-frequency words too, so click on SORT/LIMIT and change the setting for SORTING there to “Alphabetical”. Now you can search. You’ll find your results under the tab labeled FREQUENCY. But that’s not the end of the story. Unfortunately, what you will get when you search for almost any affix is a long list that needs a lot of cleaning up. There will be misspelled words, items with funky punctuation, and lots of items that just happen to end or begin with the same sequence as the affix you’re searching for, but don’t really have that affix (as was the case in our backwards word list). So you’ll need to get rid of a lot of junk before you’re left with only data on the affix your studying. To do the cleaning, you can copy your results to a spreadsheet and delete the junk, or just copy down the examples that have the affix that you’re studying, if you want a low-tech method.

Now you get to look at your data. The payoff of what may seem like a lot of grunt work is that for most English affixes you’ll come up with many more examples than you might have thought of yourself, and some of them will undoubtedly be quite interesting and surprising. And from here on, you get to hatch your own theories! Note that if you see an interesting example and you want to see the context in which it’s used, you can click on the word. An excerpt of the text in which the word is found will appear. If you want to see a bit more of the context, click on the name of the source, which can be found to the left of the excerpt. Looking at the context will help you figure out the meaning of the word and its part of speech, all of which will help you in formulating word formation rules like the ones we worked on in [Section 3.3.1](#).

A couple of final notes. First, if you want to use one of these corpora more than a few times, you’ll need to register. This is free, and takes just a couple of minutes. The website does limit the number of searches you can make per day, however, unless you purchase a subscription. If you plan to make heavy use of the corpus, you might want to check whether your university has a subscription that students can make use of. Second, the same kind of work can be done on languages other than English. Corpora exist for many languages these days. However, they are not always freely available, and the user interface will differ from one to the next.

Summary

In this chapter we have looked at a number of ways in which new words may be formed in languages. Affixed words are formed by word formation rules that make explicit the categorial, semantic, and phonological requirements of particular affixes, and specify the

categorial, semantic, and phonological properties of the resulting words. Words formed by affixation have internal structure that may be represented in the form of trees. Similarly, compound words – words composed of two or more free morphemes or bound bases – have internal structure that can be represented in trees. Compounds may be classified according to the parts of speech of the elements that make them up, or according to the semantic relations they display (attributive, coordinative, or subordinative) or according to their headedness. We have also looked at conversion, a shift in the category of a lexeme with no accompanying change in form, as well as a number of forms of word formation – coinage, blending, clipping, backformation, acronyms and initialisms, that play a minor role, at least in English. Finally, we have learned some techniques for searching for our own morphological data.

Exercises

- Divide the following words into morphemes and label each morpheme as a *prefix*, *suffix*, *free base*, or *bound base*. If you find difficulties in segmenting and labeling parts of these words, discuss what issues you find (see [Section 3.3.5](#)).
 - beginning*
reheatable
non-smoker
impurity
 - intermediate*
telephonic
overcompensation
reheatability
 - advanced*
hypoallergenic
non-morphological
unrepresentative
- On p. 41 we gave the word formation rules for *-ize*. Now consider the words below and discuss what other sorts of restrictions we would have to add to our rules for *-ize*.
catechize, evangelize, antagonize, metabolize
- Using the data below, write a word formation rule for the suffix *-able*. Consider what category it attaches to, and what part of speech the resulting words belong to. Does it seem to have any phonological or semantic restrictions? Then draw the word trees for the words *unwashable* and *rewashable*.

washable	*yawnable
dryable	*arriveable
heatable	*fallable

readable
lovable
knowable

*blinkable

4. The word *unwindable* is potentially ambiguous. What are its two possible meanings? Draw two tree structures and show which meaning goes with each structure.
5. The linguist Laurence Horn has argued (2002) that the prefix *un-* really does attach to nouns, contrary to what we said in [Section 3.3](#). He has collected such examples as *undeath*, *uncountry*, *uncopier*, *unphilosophy*, and *unpublicity*. Can you think of or find other examples where *un-* has attached to nouns? What do you think these *un-* nouns mean? (You can use a corpus or dictionary to help think of examples.)
6. In [Section 3.3](#), we discussed the meanings (or lack thereof) of bases like *-ceive*, *-mit*, and *-port*, but not the meanings of the prefixes with which they combine. Consider the prefixes *re-* and *de-* in words like *report*, *deport*, *receive*, *deceive*, *remit*, and *demit*. Do these seem to be the same prefixes as the *re-* and *de-* in *rewash*, *rewind*, *reload* or *debug*, *de-ice*, *derail*? Why or why not?
7. How many meanings can you come up with for the complex compound *miniature poodle groomer manual*? Try to draw the trees that correspond to each meaning you've come up with.
8. Classify the compounds below as attributive, coordinative, or subordinative, and as either endocentric or exocentric. Example: *book shelf* is an endocentric attributive compound; *truck driver* is an endocentric subordinate compound.

oil burner
lighthouse
blue blood
hell raiser
scholar athlete
blue-eyed
pickpocket
house-hunting
9. Many languages use compounding as a strategy for forming new words. Consider the data below and try to determine: (a) which element is the head, (b) whether the resulting compounds are endocentric or exocentric.

a. *Kannada (Dravidian)* ([Sridhar 1990](#))

a: Du-ma:tu	'speak word'	'colloquial speech'
siDi-maddu	'explode chemical'	'explosive (i.e., chemical that explodes)'
maduve a:gu	'marriage become'	'to get married'
santo:Sa paDu	'happiness feel'	'rejoice'
kittaLe haNNu	'orange fruit'	'tangerine'

b. *Maori (Austronesian)* ([Bauer 1993](#))

ipu para	'container waste'	'rubbish bin'
apuru teepu	'cushion table'	'desk pad'
wai mangu	'water black'	'ink'
whaka-koi pene	'cause.sharp pen'	'pencil sharpener'

10. Consider the following noun/verb conversion pairs in English. In each case decide whether the noun was converted from the verb or *vice versa*. Give arguments based on meaning to support your choices.

bug	to bug
kick	to kick
saddle	to saddle
paint	to paint
catch	to catch
book	to book (e.g., a table in a restaurant)

11. Take a look at the words you (and your classmates) have collected so far in your Word Logs. Can you classify them according to the means of word formation used to create them? Does any one means of word formation predominate? If so, think about why this might be.
12. Following the directions in [Section 3.8](#), use COCA to search for the suffix *-eer* (as in *auctioneer* or *mountaineer*). Make a list or a spreadsheet of the data that you find, and use those data to formulate a word formation rule for *-eer*.
13. In this chapter, we used *-tarian*, *-licious*, and *-tastic* as examples of splinters. See if you can think of other examples of splinters in English and search for novel words using those splinters in COCA.
14. Consider the semantic categories of affixes discussed in [Section 3.3.3](#). First look at the data below from the Yuman language Hualapai, spoken in northern Arizona. Assume that a process of affixation creates the pairs of verbs in this language. What semantic category would you place these affixes in? (Data from [Watahomigie, Bender, and Yamamoto 1982: 280](#).)

d-akk	'to throw toward the speaker'	d-amk	'to throw from the speaker'
e:kk	'to give/receive (toward me)'	e:mk	'to send'
ha:kk	'to look this way'	ha:mk	'to look over that way'
jjiyu:kk	'to send one person/animal toward the speaker'	jjiyu:mk	'to send one person/animal away'
'u:kk	'to come and see'	'u:mk	'to go and see'
vo:kk	'to come home'	vo:mk	'to go home'

4 Productivity and Creativity

CHAPTER OUTLINE

KEY TERMS

productivity
transparency
lexicalization
compositionality
frequency
creativity
hapax
legomenon

In this chapter you will learn about productivity – the extent to which word formation rules can give rise to new words.

- ◆ We will consider what factors contribute to productivity, what restricts the productivity of word formation processes, and how we can measure productivity.
- ◆ We will look at how the productivity of a word formation process can change over time.
- ◆ And we will consider how speakers of a language can use even unproductive word formation processes to create new words for humorous or playful effects.

4.1 Introduction

Consider the examples in (1):

- | | | | |
|-----|----|--------|-----------|
| (1) | a. | warm | warmth |
| | | true | truth |
| | b. | modern | modernity |
| | | pure | purity |
| | c. | happy | happiness |
| | | dark | darkness |

In each case, we have adjectives and nouns that are derived from them (all cases of transposition, by the way). As a first pass, we might hypothesize the three rules of lexeme formation in (2):

- (2)
- a. Rule for *-th*: *-th* attaches to adjectives, and creates nouns. For a base meaning 'X', the derived noun means 'the state of being X'.
 - b. Rule for *-ity*: *-ity* attaches to adjectives, and creates nouns. For a base meaning 'X', the derived noun means 'the state of being X'.
 - c. Rule for *-ness*: *-ness* attaches to adjectives, and creates nouns. For a base meaning 'X', the derived noun means 'the state of being X'.

Now consider the list of adjectives in (3). If you had to make a noun from each of these, which of the three suffixes would you choose (note that you might be able to use more than one in some cases)?

- (3)
- lovely
 - cool
 - crude
 - evil
 - googleable
 - rustic
 - musty
 - inconsequential
 - feline
 - toxic
 - bovine

Chances are that there are some of these words that you would choose to use *-ity* with (I choose *crude*, *toxic*, *googleable*, *rustic*, *inconsequential*, maybe *feline*), and others that you would use *-ness* with (for me, *lovely*, *cool*, *evil*, *musty*, probably *bovine*). Your choices might be slightly different from mine, but I'd be willing to predict that you didn't choose to use *-th* with any of these adjectives.

What does this mean? In some cases, we can look at words, decide that they are complex, and isolate particular affixes. But when it comes to using those affixes to create new lexemes, we have the sense that they are no longer part of our active repertoire for forming new words. We have no trouble using other affixes, however, even if we've never seen them on particular bases; for example, you may never have seen a noun

form of the word *bovine*, but you have no trouble forming the word *bovineness* (or maybe *bovinity*, or maybe even both). Processes of lexeme formation that can be used by native speakers to form new lexemes are called **productive**. Those that can no longer be used by native speakers are unproductive; so although we might recognize the *-th* in *warmth* as a suffix, we never make use of it in making new words. The suffixes *-ity* and *-ness*, on the other hand, can still be used, although perhaps not to the same degree. Most morphologists agree that **productivity** is not an all-or-nothing matter. Some processes of lexeme formation, like affixation of *-th*, are truly unproductive, but for those processes that are productive, we have the sense that some are more productive than others. In this chapter we will explore in some detail what we mean by productivity, and look at a number of factors that contribute to productivity. We will also look at several ways in which productivity can be measured.

Challenge

Consider the words below:

sparkle, rattle, drizzle, rumble, shuffle, ripple, crumble, wiggle,
twinkle, tremble, dribble, rustle, nibble

All of these words end in *-le*. In what ways do you find these words similar? Given the similarities that you've noticed, do you think that *-le* should be considered a suffix in contemporary English? If so, what would you say about its productivity?

4.2 Factors Contributing to Productivity

A number of factors contribute to the degree to which we can use morphological processes to create new lexemes (see [Figure 4.1](#)).

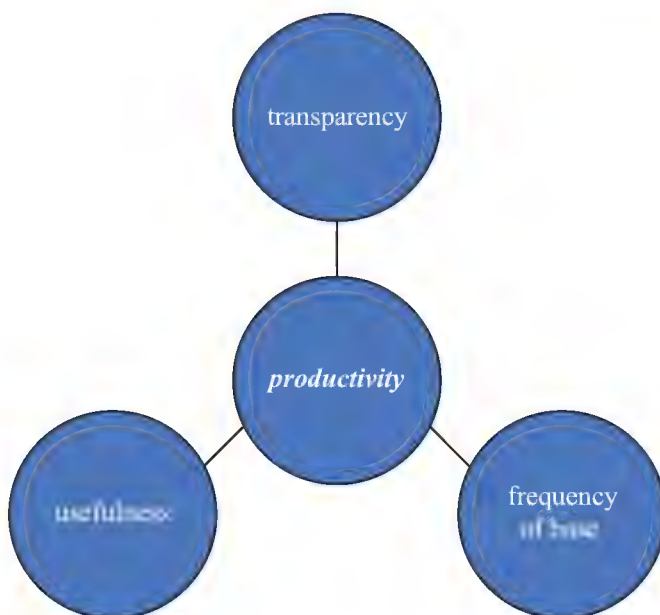


FIGURE 4.1
Factors contributing to
productivity

One factor is what is called **transparency**. Words formed with **transparent processes** can be easily segmented, such that there is a one-to-one correspondence between form and meaning. In other words, when we attach an affix to a base, the phonological form (the pronunciation) of both morphemes stays the same, and the meaning of the derived word is exactly what we would expect by adding the meaning of the affix to that of the base. Let's look further at the case of *-ness* and *-ity*, this time considering the additional examples in (4):

- | | | | |
|-----|----|-------------|----------------|
| (4) | a. | candid | candidness |
| | | pink | pinkness |
| | | hardy | hardiness |
| | | common | commonness |
| | | ticklish | ticklishness |
| | | cunning | cunningness |
| | | horrible | horribleness |
| | | pure | pureness |
| | | odd | oddness |
| | b. | crude | crudity |
| | | odd | oddity |
| | | pure | purity |
| | | dense | density |
| | | rustic | rusticity |
| | | timid | timidity |
| | | grammatical | grammaticality |
| | | local | locality |
| | | available | availability |
| | | senile | senility |

In all the *-ness* examples in (4a), it is easy to divide the complex words into base and suffix. The base is always pronounced in the derived word as it is in isolation. And the suffix always creates a noun meaning 'state of being X', where X is the adjective base. Words formed with *-ness* are perfectly transparent. The suffix *-ity* is somewhat less transparent. Although you don't see this when words are written in English orthography, when you pronounce them, you see that *-ity* often has the effect of changing the phonological form of its base – sometimes its stress pattern, and sometimes both stress and phonological segments in the base. So *timid* in isolation is pronounced with stress on the first syllable (*tímid*), but when *-ity* is added, stress shifts to its second syllable (*timíidity*). And with the base *rustic*, in addition to a shift in stress from first to second syllable, the final [k] of the base becomes [s] when *-ity* is added. Further, some of the words formed with *-ity* have meanings that cannot be arrived at by combining the meaning of the base with that of the suffix. An *oddity*, for example, is not merely 'the state of being odd' (we would probably prefer the word *oddness* for that meaning), but a person or thing that is odd. And a *locality* is not 'the state of being local', but a place or area. Finally, consider the examples in (5):

- (5) verity
dexterity
authority

In the first two examples, *-ity* occurs on bound bases *ver*, *dexter*. In the third, it's not clear exactly how to analyze the derived word. Although it appears to be a combination of *author* and *-ity*, there are two problems with this analysis. First, as a free base *author* is a noun, and *-ity* typically attaches to adjectives, rather than nouns. And second, it's not clear what the independent meaning of the base is; certainly the meaning 'professional writer' does not seem to be part of the meaning of *authority*. We never find *-ness*, however, on bound bases. All of this adds up to a conclusion that the suffixation of *-ness* is a much more transparent process than the suffixation of *-ity*, and this in turn suggests that *-ness* is a more productive affix than *-ity*.

Hand in hand with the notion of transparency comes the related notion of **lexicalization**. When derived words take on meanings that are not transparent – that cannot be made up of the sum of their parts – we say that the meaning of the word has become lexicalized. Meanings of complex words that are predictable as the sum of their parts are said to be **compositional**. Lexicalized words have meanings that are **non-compositional**. So the words *oddity* and *locality* that we looked at above have developed lexicalized or non-compositional meanings. Sometimes the meanings of derived words have drifted so far from their compositional meanings that it's quite difficult to imagine the compositional meaning for them. Consider, for example, the word *transmission*, which denotes a part of a car. It takes a bit of thought to realize that the car part in question is so called because it transmits power from the engine to the wheels.

Transparency is not the only factor that contributes to productivity. Another factor that is important is what we might call **frequency of base type**. By this, I mean the number of different bases that might be available for affixes to attach to, thus resulting in new words. If an affix attaches only to a limited range of bases, it has less possibility of giving rise to lots of new words, and it will therefore be less productive. Consider, for example, the suffix *-esque* in English, which means something like 'having the style of' (Marchand 1969: 286).

Challenge

Do a COCA search for the suffix *-esque* following the instructions in Section 3.8. What kinds of bases does *-esque* attach to? Look at the category (part of speech) of the bases, the semantic types of those bases, and the phonological properties of those bases.

You'll probably find that *-esque* attaches to nouns, but mostly to concrete ones (*mountainesque*), and in fact, most often to proper names

(*Kafkaesque*, *Bidenesque*). Indeed, although it attaches pretty freely to names, it seems most comfortable on names that have at least two syllables (?*Trumpesque*, ?*Penceesque*). Compared to a suffix that could attach to any noun at all, *-esque* would be less productive.

The final factor that contributes to productivity is what we might call **usefulness**. A process of lexeme formation is useful to the extent that speakers of a language need new words of a particular sort. It's always useful, for example, to be able to form a noun meaning 'the state of being X' from an adjective, whatever X means, so both *-ness* and *-ity* are highly useful affixes. On the other hand, consider the suffix *-ess* in English. It used to be useful to be able to coin words referring to jobs performed by women or positions held by women (*stewardess*, *murderess*, *authoress*). But with the rise of feminism and efforts to promote gender-neutral language, such words have fallen into disuse, and the need for new words using this suffix has almost died out. Consequently the affix has become far less productive.

4.3 Restrictions on Productivity

As we saw above, the more limitations there are on the bases available to a lexeme formation process, the less productive it will be. In this section, we will explore different kinds of restrictions that may apply to lexeme formation processes.

We have actually looked at some such restrictions in [Chapter 3](#) ([Section 3.2](#)), when we learned how to write word formation rules. We learned that there could be different sorts of restrictions on what sorts of base an affix might attach to, including:

- *categorial restrictions*: Almost all affixes are restricted to bases of specific categories. For example, *-ity* attaches to adjectives, *-ize* attaches to nouns and adjectives, or *un-* attaches to adjectives or verbs.
- *phonological restrictions*: Sometimes affixes will attach only to bases that fit certain phonological patterns. For example, *-ize* prefers nouns and adjectives that consist of two or more syllables, where the final syllable does not bear primary stress. The suffix *-en*, which forms verbs from adjectives, attaches only to bases that end in obstruents (stops, fricatives, and affricates). So we can get *darken*, *brighten*, and *deafen* but **slimmen* and **tallen*, which end in sonorant consonants, are impossible.
- *the meaning of the base*: For example, negative *un-* prefers bases that are not themselves negative in meaning. We find *unlovely* but not **unugly*, *unhappy* but not **unsad*.

To these sorts of restrictions we might add:

- *etymological restrictions*: Some affixes are restricted to particular subclasses of bases. For example, there are affixes in English that prefer

to attach to bases that are native Germanic – for example the suffix *-en* that forms adjectives from nouns (*wooden*, *waxen* but not **met-alen* or **carbonen*). On the other hand, another suffix *-ic* that forms adjectives from nouns (*parasitic*, *dramatic*) will not attach to native bases, only to bases that are borrowed into English from French or Latin.

- *syntactic restrictions*: Sometimes affixes are sensitive to syntactic properties of their bases. For example, the suffix *-able* generally prefers to attach to transitive verbs (verbs that have a subject and an object), specifically verbs that can be passivized. So from the transitive verb *love* we can get *lovable*, but from the intransitive verb *snore* there is no **snorable*.
- *pragmatic restrictions*: Bauer (2001: 135) gives the following example. In Dyirbal, there is a suffix *-ginay* that means ‘covered with’. Although there might conceivably be a use for a word meaning something like ‘covered with honey’, in fact, the suffix occurs in Dyirbal only on bases that denote things that are “dirty or unpleasant” (Dixon 1972: 223), like *gunaginay*, which means ‘covered with feces’. What’s considered dirty or unpleasant might to some extent be a function of cultural beliefs.

We might expect there to be an inverse correlation between the number of restrictions and the productivity of a lexeme formation process: the more restrictions apply, the fewer bases it will have available to it, and the fewer words it will be able to derive.

The restrictions above pertain to inputs to lexeme formation rules. But it’s also possible for there to be restrictions specifically on the output of rules. For example, certain sorts of complex words can be restricted in register. Baayen (1989: 24–5) notes that the suffix *-erd* in Dutch forms “jocular and often slightly pejorative personal names.” For example, from the adjective *bang* ‘afraid’ we get *bangerd* ‘fraidy-cat’ and from *dik* ‘fat’ we get *dikkerd* ‘fatty’. Baayen points out that although there are a lot of adjectives that might give rise to pejorative names for people, words formed with the suffix are confined to use in spoken, as opposed to written, language and therefore the output of this lexeme formation process is restricted.

4.4 Ways of Measuring Productivity

We have seen that the productivity of lexeme formation processes depends on a variety of factors, including restrictions on possible bases, usefulness of the words formed, and the transparency of the process. Looking at these factors can give us some sense of how productive a process might be, but can we do better and actually measure productivity? Is it possible to compare the productivity of different processes? If so, how might we go about making such measurements?

Challenge

One conceivable way of measuring the productivity of a lexeme formation process might be to count up all the items formed with that process that can be found in a good dictionary. Most morphologists think that this is not a good way of measuring productivity. List as many reasons as you can why they should think so.

It's not hard to think of reasons why counting items in a dictionary wouldn't be an accurate way of estimating productivity. For one thing, counting items that are already in the dictionary doesn't really tell us anything about how many new words might be created with a lexeme formation process, and it's the possibility of creating new forms that's most important in making processes productive. Further, the most productive of lexeme formation processes are ones that are phonologically and semantically transparent. If the words resulting from these processes are perfectly transparent in meaning, then it's unlikely that dictionaries will need to record them! On the other hand, less productive processes, as we've seen, frequently have outputs that are less transparent (more lexicalized), and therefore have more need to be listed in the dictionary. So simple counting might give a paradoxical result: less productive processes would be represented by more entries in the dictionary than more productive processes!

Morphologists have therefore tried hard to come up with other ways of measuring productivity. One suggestion (Aronoff 1976) was to make a ratio of the number of actual words formed with an affix to the number of bases to which that affix could potentially attach.

Challenge

Consider the suffix *-esque* that we mentioned in Section 4.2 above. Based on the data that you gathered yourself, do you see any problem in counting the number of bases to which *-esque* could potentially attach?

Most morphologists would say that we might be able to count the words we find in a given dictionary with a particular affix, but would this number be equivalent to the number of words in any speaker's mental lexicon? How could we even guess at this? And further, how do we count potential bases? In the case of *-esque*, this would require knowing how many names the suffix might attach to, which is, of course, not possible!

A somewhat more sophisticated – but still not perfect – measure of productivity proposed by Baayen (1989) capitalizes on what we know about the token frequency of derived words. Remember from Chapter 1 the distinction between types and tokens: if we're counting types in a corpus or language sample we look for each different word and count it once, no matter

how many times it appears, but if we're looking at tokens we count up all the separate occurrences of that word in a particular corpus. The number of separate occurrences of a word in the corpus is the token frequency of that word.

An important observation that has been made about lexeme formation processes is that the less productive they are, the less transparent the words formed by those processes, and the less transparent the words, the higher their mean token frequency in a corpus. In other words, words formed with less productive suffixes are often more lexicalized in meaning and will often display many tokens in a corpus. The more productive a process is, the more new words it will give rise to and the more chance that these items will occur in a corpus with a very low token frequency, sometimes only once. Remember that in [Section 3.8](#) I suggested that we might be more interested in derived words that occur with a very low frequency, rather than ones that occur with a high frequency in a corpus? This is why: we can make use of this observation by counting up all tokens of all words formed with a particular affix, and then seeing how many of those words occur only once in the corpus (a type with token frequency of one in a corpus is called a **hapax legomenon** or sometimes just a **hapax**). The ratio of hapaxes to all tokens tells us something about the probability of finding new forms using that affix: the higher the probability of low-frequency items, the likelier those items are to be new formations. [Baayen \(1989\)](#) gives us this formula:

(6) *Baayen's productivity formula*

$$P = \frac{n_1}{N}$$

where N equals the total number of tokens and n_1 the number of hapaxes.

For example, using my own data from COCA, I count 5,738 total tokens for words using the suffix *-esque*, of which 665 types occur only once. This gives a P score of .12 for *-esque*. To know exactly what this number means, we need to look at it relative to other suffixes in contemporary English. Take, for example, the suffix *-dom*, which attaches primarily to nouns and forms nouns meaning 'the state or quality of Noun', as in *chiefdom*, *fandom*, and *stardom*. Again, using my own data from COCA, I count 64,345 total tokens and 257 hapaxes. That gives a P score of .004. You can see that by comparison, *-esque* appears to be far more productive than *-dom*.

4.5 Historical Changes in Productivity

It should not come as a surprise at this point that lexeme formation processes may change their degree of productivity over time. Particular prefixes or suffixes can come into vogue and flourish for a few decades or centuries and then fade out of existence. The words formed with

them may persist for a long time, but we won't find any new ones being formed. In this section we will see how it's now possible to track the productivity of affixes in English.

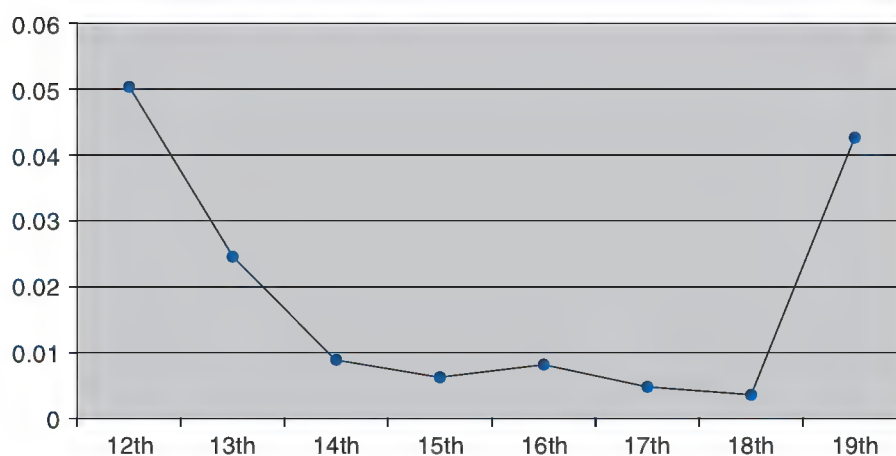
How can we measure changes in productivity? Until recently, the best way that we had to gauge the historical productivity of a prefix or suffix was by using dictionaries. For example, the *Oxford English Dictionary* gives us information about when particular words are cited for the very first time in English, and from looking at those first citations we can see something about how many new words have been formed in a given century with a particular affix. If we know how many first citations there are for words with a given affix in a specific century, and if we also know the total number of citations there are in the *OED* for each century (not every century has the same number of citations), we can calculate what percentage of them are first citations with that affix.

Let's see how this works with the suffix *-dom*. For example, if the *OED* gives 28,698 citations dating from the thirteenth century, and 7 of them are the first citations of words with *-dom*, then the *-dom* citations represent 0.0243% of all those citations. I've calculated these percentages for each century, and then plotted them on a graph, as you can see in Figure 4.2.

The suffix *-dom* is a very old one, going back to the beginnings of English, and indeed further back into the Germanic branch of the Indo-European family, from which English descends.¹ We can see that after an initially very productive period in the twelfth century, *-dom* seems to have dropped off in productivity from the fourteenth through the eighteenth centuries. But its productivity rises again in the nineteenth century.

These days, if we don't want to trace the productivity of an affix over its whole history, but only over the last two or so centuries, we can make use of Baayen's P formula. To do this, we can use the Corpus of Historical

FIGURE 4.2
Comparative percentages
of first citations of *-dom*
per century



1. Figure 4.2 starts with the twelfth century because that is the first century for which we have the number of citations in the *OED*. Many thanks to Charlotte Brewer of Oxford University for sharing these data with me.

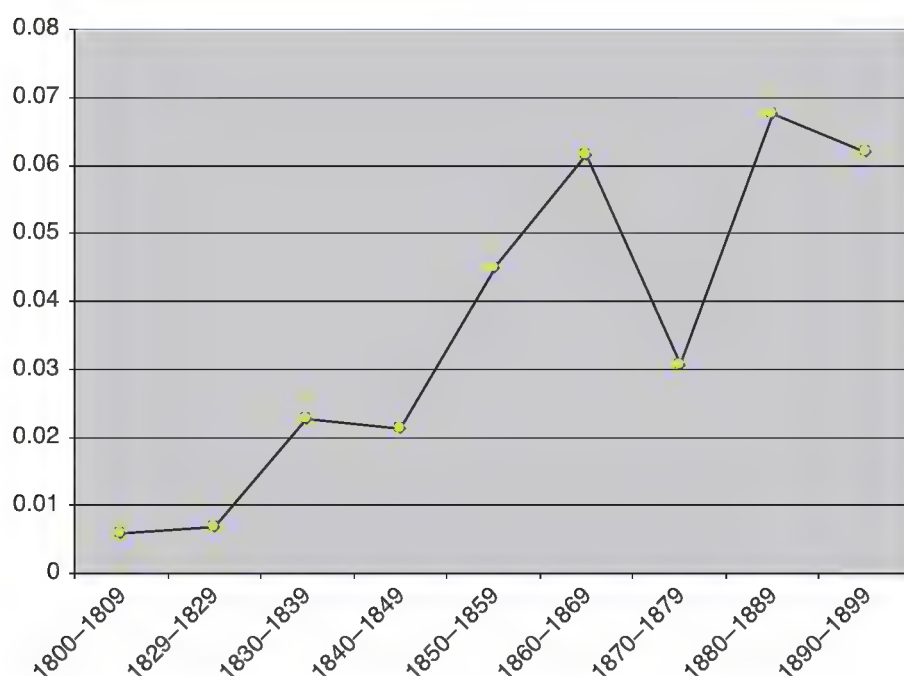


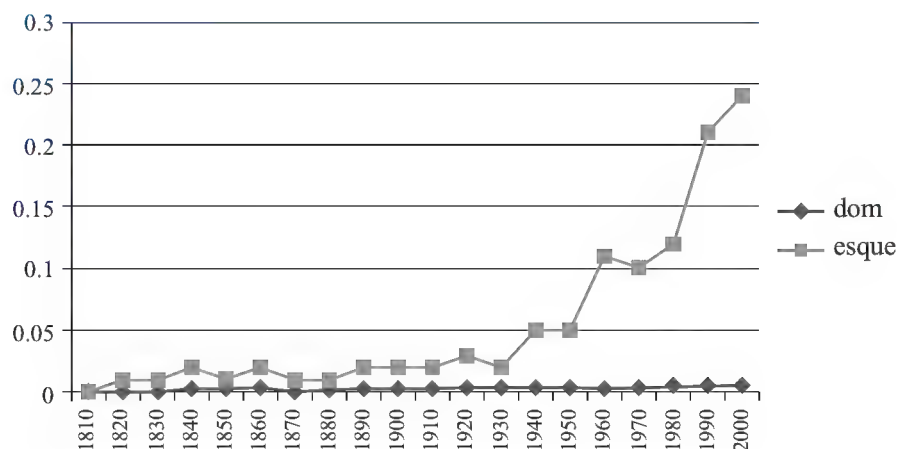
FIGURE 4.3
Productivity of *-dom*
using Baayen's P
measure and data from
COHA

American English (COHA). COHA breaks the data down by decade, starting with 1810 and ending with 2000. As we did with COCA, we can use COHA to see how many tokens overall there are with a particular affix and how many of those tokens are hapaxes. And from these two numbers we can calculate the P score for that affix decade by decade. As you can see from Figure 4.3, the P score for *-dom* generally rises from the beginning of the nineteenth century to the end of the twentieth century, although it's not a completely smooth rise – there's a mysterious dip in the 1860s. We can only wonder about the decline in the popularity of the suffix during the decade from 1860 to 1870!²

Exactly why the productivity of this suffix should start to rise after centuries of minimal productivity is unclear. Wentworth (1941) notices the same trend that I've shown here, and points out that particular nineteenth-century authors seem especially prone to coin new words with the suffix: Thomas Carlyle and William Makepeace Thackeray in Britain, Mark Twain and Sinclair Lewis in the United States. But whether they are the cause of the rise or a reflection of something that was happening in the language at large is impossible to say. Bauer (2001) points out that in the nineteenth century, the kind of bases available to the suffix *-dom* seemed to expand drastically. Where it was confined for many centuries to bases referring to important types of people (*lord*, *king*, *master*, *pope*, *earl*, but also *martyr* and *witch*) or a few adjectives (*wise*, *rich*, *free*), in the nineteenth century it began to appear with more frequency on names for animals (*puppy*, *dog*, *butterfly*, *centaur*) and a wide variety of common nouns (*school*, *twaddle*, *leaf*, *magazine*, *jelly*, *cotton*, *fossil*). But why, exactly, its range of bases expanded at this point is still a mystery.

2. The careful reader will note that the P score for *-dom* in the last two decades of the twentieth century looks higher than the figure I calculated for *-dom* using COCA above. The reason for the discrepancy is that COHA uses a different sample of texts than COCA. The overall number of words per decade is smaller, and all excerpts are from written English, so finding a small discrepancy in P figures should not be especially troubling.

FIGURE 4.4
Comparative productivity
of *-esque* and *-dom* using
data from COHA



Note that we can also use COHA to compare the relative productivity of affixes over the last two centuries. Figure 4.4 illustrates this with the suffixes *-dom* and *-esque*. Remember that we calculated P for both suffixes on the basis of the contemporary data from COCA and found that *-esque* appeared to be much more productive than *-dom*. The data from COHA enriches our understanding of the relative productivity of these two suffixes. Consider Figure 4.4. You can see from the graph that compared to *-dom*, the suffix *-esque* not only is much more productive overall, but that the trajectory of its productivity is different. The suffix *-esque* seems to have taken off in productivity starting in the 1940s, and has risen dramatically since 1980, whereas the rise in productivity of *-dom* is comparatively modest.

Challenge

Try your hand at tracing the historical productivity of the prefixes *mini-* and *nano-* in English. Can you pinpoint when each one started to gain in productivity?

4.6 Productivity versus Creativity

Some morphologists make a distinction between morphological productivity and morphological **creativity**. When processes of lexeme formation are truly productive, we use them to create new words without noticing that we do so. Similarly, when hearers are exposed to a productively formed complex word, they understand it, but usually don't note that it's a new word (at least for them). This is not to say that speakers and hearers *never* notice productively formed new words, just that often such words slip by without notice. Morphological creativity, in contrast, is the domain of unproductive processes like suffixation of *-th* or marginal lexeme formation processes like blending or backformation. It occurs when speakers use such processes consciously to form new

words, often to be humorous or playful or to draw attention to those words for other reasons.

For example, speakers might use the unproductive suffix *-th* to form an adjective like *coolth* (in contrast to *warmth*), consciously trying to be clever or witty. Another example might be the suffix *-some* that occurs in English in words like *twosome*, *threesome*, and *foursome*. Theoretically this suffix might be infinitely productive because its bases are cardinal numbers. But it's really only attached to the numbers *two* through *four* or *five*. We would probably only coin a new word like *seventeensome* if we were trying to be funny. Such a use would be creative, rather than a productive use of this lexeme formation process.

Let's now look more closely at the case of blending in English. **Blending**, as we saw in [Chapter 3](#), is the creation of new words by putting together parts of words that are not themselves morphemes. Relatively few blended words have become lexicalized words in English (*brunch*, *smog*), but the technique is frequently used for coining words by advertizers and the media, precisely because such words are noticeable. McDonald's, for example, creates a word like *menunaire* from *menu* and *millionaire* to catch your eye (or ear), and make you pay attention to their pitch.

Many blends can be culled from social media and the popular press. For example, in the first few months of the Covid-19 pandemic, you might have seen the following words in newspapers or on various websites covering the pandemic:

(7) maskne	blend of <i>mask</i> and <i>acne</i> . 'Blemishes on your face from wearing a mask.'
quarantini	blend of <i>quarantine</i> and <i>martini</i> . 'Cocktail made during the quarantine.'
quarbaking	blend of <i>quarantine</i> and <i>baking</i> . 'Excessive baking done out of boredom during the quarantine.'
zoombies	blend of <i>zoom</i> and <i>zombies</i> . 'People who turn into zombies from too many Zoom meetings.'
covidiot	blend of <i>Covid</i> and <i>idiot</i> . 'Person who acts foolishly in ways related to the pandemic.'
plandemic	blend of <i>plan</i> and <i>pandemic</i> . 'Conspiracy theory that the Covid pandemic was planned for some nefarious reason.'
coronacoaster	blend of <i>corona</i> and <i>rollercoaster</i> . 'The emotional instability that is a consequence of Covid anxiety.'
maskhole	blend of <i>mask</i> and <i>asshole</i> . 'Disparaging term for a person who refuses to wear a mask.'

All of these words found their way from their original print sources to the NOW corpus. We might speculate that such words are intended to catch the reader's eye and therefore make for lively reading.

As [Bauer \(2001\)](#) points out, however, it is not always possible to draw a sharp line between productivity and creativity. Take, for example, the diminutive suffix *-let* in English (*booklet*, *wavelet*, *eyelet*). A look at the *Oxford*

English Dictionary and COHA suggests that this suffix enjoyed at best a small degree of productivity during the nineteenth and twentieth centuries. COHA gives us a handful of apparently novel forms, that is, ones that occur only once in the corpus and are not captured in the OED: *peaklet*, *gravelet*, *adlet*, *auntlet*, *boatlet*, *commissionlet*, *clifflet*, *cluster-bomblet*, *choplet*, *czarlet*, *dictatorlet*, *flowerlet*, *flylet*, *footnotelet*, *faunlet*, *dudelet*, *statelet*, *porchlet*, *snifflet*, *squablet*, *plotchlet*, *sky-let*, *shoelet*. Most of these forms are used in contexts which suggest not just small size, but also disparagement and often irony. To these I can add a further example that comes from a biography of Julia Child (great TV chef and cookbook author), in which her husband refers to her as his *wifelet* – surely meant to be funny, as Julia was over six feet tall!³ This last example was clearly meant to draw attention to itself, but are all these *-let* formations self-conscious enough that they should be seen as creative formations, or rather is *-let* just a minimally productive suffix that's taken on a special ironic or pejorative flavor in English? Perhaps there is no good way of answering this question, but looking at examples in the corpora gives us a new way of considering the issue for ourselves.

Summary

In this chapter we have explored the notion of productivity – the extent to which lexeme formation processes can be used to create new words. We have seen that several factors contribute to productivity: the phonological and semantic transparency of the process, the size of the pool of bases it can apply to, and its usefulness. We have seen as well that there can be different sorts of restrictions on lexeme formation processes that result in a decrease in productivity. Among these are categorial, phonological, semantic, syntactic, etymological, and pragmatic restrictions. We have looked as well at ways in which we can measure productivity. Finally, we have seen that even unproductive and marginal processes can still give rise to occasional new formations, a phenomenon that we called creativity, but that discerning the line between productivity and creativity may be a subjective matter.

Exercises

- Which of the following derived words with the suffix *-ity* have lexicalized (non-compositional) meanings. Hint: some have both. Fill in the grid below:

	Compositional? Yes/No	Compositional meaning	Non-compositional meaning
a. curiosity			
b. solidity			

3. Laura Shapiro, *Julia Child*. Viking, 2007.

c. publicity			
d. sexuality			
e. visibility			
f. facility			

2. Consider the examples in (a)–(c) below. Each set involves a lexeme formation process that takes nouns as base and produces adjectives. On the basis of these examples, compare the three lexeme formation processes in terms of their transparency. Remember that transparency involves both compositionality of meaning and the phonological stability of the base (i.e., the base is pronounced the same way in isolation and in the derived word):
 - a. *-ish* girlish
kittenish
sheepish
loutish
babyish
 - b. *-ic* cyclic
metallic
economic
totemic
organic
 - c. *-al* herbal
global
homicidal
glacial
clinical
3. In this chapter, we have looked exclusively at productivity as it concerns derivational processes. We can, however, also compare the productivity of various types of compounding. English has compounds that consist of two nouns (*dog bed*, *windmill*), two adjectives (*bittersweet*, *blue-green*), and two verbs (*blow dry*, *stir-fry*). Are all three types of compounding equally productive? (Hint: one way to start is by thinking of examples of NN, AA, and VV, and seeing which type gives you the most difficulty.) Give as much evidence as you can for your answer.
4. Look at the words you've collected in your Word Log. How many of them are formed with affixes? Which affixes? How many are formed by compounding, conversion, or blending? What does this tell you about the productivity of various processes in present-day English?
5. The graph in [Figure 4.5](#) shows percentages of first citations in the *OED* with the suffixes *-dom*, *-esque*, *-ship*, *-let*, and *-hood*. Make some observations on the patterns that you observe in the graph. How good a view of the comparative productivity of these suffixes do you think this chart gives? Take into consideration what we have said in this chapter about basing estimates of productivity on material in a dictionary.

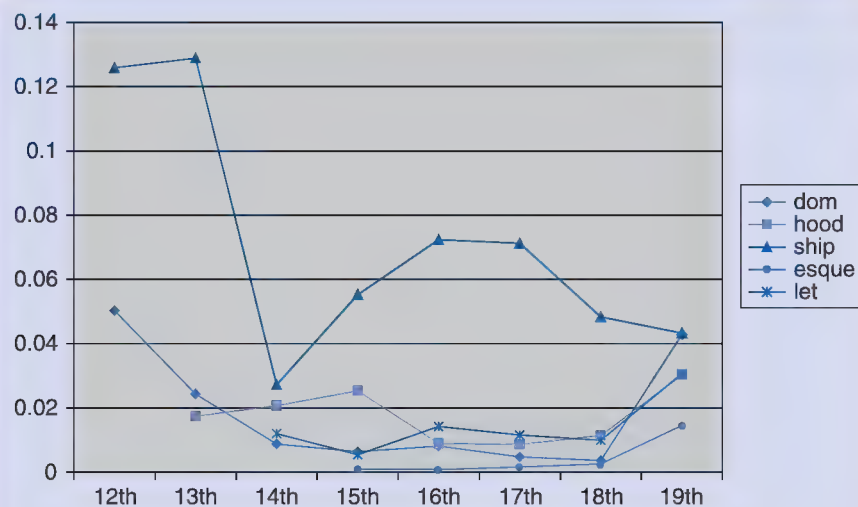


FIGURE 4.5

Percentages of first citations in the OED of suffixes *-dom*, *-esque*, *-ship*, *-let*, *-hood*.

6. Using COHA search for words ending in the suffix *-eer* (as in *charioteer* or *mountaineer*). You'll first need to clean the data, as you did for the exercises in [Chapter 3](#). Then look for all the *-eer* words which occur with a frequency of 1. Check which decade each word appeared in and see if you can make a graph showing how many hapaxes occurred in each decade. (You don't need to do a P calculation – just count items with a frequency of 1 in each decade.) Can you see any trends in the productivity of the suffix *-eer*?
7. Do the same for the prefix *mega-*. Discuss the trends that you find.
8. Go back to the splinters you found for exercise 13 in [Chapter 3](#). Try to determine when each one began to give rise to new forms in English. Do you see any trends in the formation of splinters?

5 Lexeme Formation: Further Afield

CHAPTER OUTLINE

KEY TERMS

infix

circumfix

parasyntesis

internal stem change

ablaut

consonant mutation

templatic morphology

reduplication

subtractive morphology

In this chapter you will begin to look in detail at data from a wide range of languages and learn about kinds of morphological processes that are uncommon or don't exist at all in English and other widely studied languages.

- ◆ We will look at kinds of word formation that may be new to you: infixation, circumfixation, ablaut, umlaut, and consonant mutation, reduplication, and templatic morphology.
- ◆ And you will get further practice in morphological analysis.

5.1 Introduction

So far we have looked at the way lexeme formation works in English and a few other well-studied languages. Of course, morphologists are interested in how word formation (and inflection, which we will get to in [Chapter 6](#)) works across all the languages of the world. Cross-linguistic investigation allows us to see the ways in which English word formation is typical and the ways in which it is unusual. From our perspective, English morphology might seem utterly typical, but it's only by comparison with other languages that we can begin to gain a wider perspective. What is typical in English might very well turn out to be atypical in the context of other languages.

In this chapter we will first learn how to work with data from a wider range of languages and then move on to kinds of word formation that are rare or completely non-existent in English. Just to give you a taste of what's to come, take a look at the data in (1):¹

- (1) a. *Tagalog (Austronesian)* ([Schachter and Otanes 1972: 356](#))
- | | | | |
|-------|--------------|---------|--------------------|
| ganda | 'beauty' | gumanda | 'become beautiful' |
| hirap | 'difficulty' | humirap | 'become difficult' |
- b. *Manchu (Tungusic)* ([Haenisch 1961: 34](#))
- | | | | |
|-------|----------|-------|----------|
| haha | 'man' | hehe | 'woman' |
| ama | 'father' | eme | 'mother' |
| amila | 'cock' | emile | 'hen' |
- c. *Samoan (Austronesian)* ([Mosel and Hovdhaugen 1992: 227](#))
- | | | | |
|------|----------|----------|-------------------|
| a'a | 'kick' | a'aa'a | 'kick repeatedly' |
| 'etu | 'limp' | 'etu'etu | 'limp repeatedly' |
| fo'i | 'return' | fo'ifo'i | 'keep going back' |

These examples should look quite different from the kinds of morphology that we've concentrated on so far: prefixation, suffixation, compounding, and conversion. In (1a), it looks like a morpheme has been inserted right into a base to form a verb. In (1b), vowels have changed to form the female correlates of male nouns, and in (1c), segments of the base are repeated to form what's called the frequentative form of the verb (for a verb meaning X, this form means 'X repeatedly'). Prefixation, suffixation, compounding, and conversion may be the main ways of forming new words in English and many other languages, but there's a much wider world out there, and there are types of morphology that do not figure in English at all, or figure only in the most minor ways.

In this chapter we'll expand our horizons by surveying a number of morphological processes that we have not yet encountered: different kinds of affixes, internal stem changes to consonants and vowels, reduplication, and templatic morphology. Our concentration will be on

1. Where possible, at the first mention of a language I will supply its genetic affiliation, as cited in Ethnologue. When citing data, I will use the name for the language that is used in the source from which I've taken the data, even though there might be alternative names for the language that are currently preferred by speakers or linguists working on the language. Information about alternative names, where the language is spoken, and numbers of speakers can be found on the Ethnologue website (www.ethnologue.com/).

the structural aspects of morphology – the kinds of rules that languages can make use of to form new words – as opposed to the semantic or grammatical aspects. Our aim here is to characterize a sort of universal toolbox of rules which languages may make use of in word formation.

5.2 How To: Morphological Analysis

Before we look at different kinds of word formation, student morphologists need to be comfortable working with data from a wide range of languages. When you look at English, you may have an intuitive sense of how to divide words into morphemes. But you might not have the same sense when you encounter a language you know nothing about. Surprisingly, with a few analytical tools at your disposal it's possible to learn quite a lot about how words are formed in a language that's new to you. In this section I'll take you through the steps of developing a working hypothesis about how the morphology works in a language you've probably never looked at before.

How do linguists go about deciding what words are complex in a language they know nothing about, what sorts of processes are involved in creating complex words, and how to analyze individual words? Consider the words in (2), from the language Dyirbal, a language of the Pama-Nyungan family; the data are taken from [Dixon \(1972: 222–33\)](#):

- | | | |
|--------|-----------------|--------------------------|
| (2) a. | nalŋgaŋunu | 'from a boy' |
| b. | yaɾaŋaɾu | 'like a man' |
| c. | gugulaŋaɾu | 'like a platypus' |
| d. | banabaɬun | 'proper water' |
| e. | waŋalɬaɬun | 'proper boomerang' |
| f. | yaɾabaɬun | 'proper man' |
| g. | yaɾagabun | 'another man' |
| h. | yaɾaɬaran | 'two men' |
| i. | baŋguyɬaran | 'two frogs' |
| j. | yugubila | 'with a stick' |
| k. | waŋalɬaranbila | 'with two boomerangs' |
| l. | miɬagabunŋunu | 'from another camp' |
| m. | gugulabaɬunŋaru | 'like a proper platypus' |
| n. | yaɾagabunɬaran | 'two other men' |

Just by looking at the Dyirbal words (2a) and (2b) and their glosses, you really can't tell anything. They might be simple or complex, but there's no way of knowing, because there are no parts of the two words that seem to overlap. But as soon as you look at example (2c) and its gloss, you will notice some overlap with (2b). Both examples share the gloss 'like a', and both have some characters at the end that overlap (*aŋaɾu*). So you might make a tentative hypothesis that these words are complex, and that they can be broken down into two morphemes, *yaɾ* + *aŋaɾu* and *gugul* + *aŋaɾu*, respectively. You might also hypothesize that *yaɾ* means 'man', *gugul* means 'platypus', and *aŋaɾu* means 'like a'. This is a good first guess, but you should always be prepared to revise your analysis as you look at more data.

If you then move on and look at examples (2d–f), you’ll notice that they all share part of their meanings (‘proper’), and the end of each word has the sequence *baɖun*. It’s therefore reasonable to make the hypothesis that *baɖun* means ‘proper’, and that what’s left over means ‘water’ in (2d), ‘boomerang’ in (2e), and ‘man’ in (2f). But now, we need to look back at our analysis of (2b), because our first hypothesis was that *yaɾ* meant ‘man’, and what’s left over in (2f) is not *yaɾ* but *yaɾa*. We therefore need to go back and revise our analysis of examples (2b) and (2c) to be consistent with what we’ve learned from examples (2d–f). This means that (2b) should be divided into *yaɾa+ŋaɾu* and (2c) should be divided into *gugula+ŋaɾu*. What we’ve discovered so far is summarized in (3):

(3)	<i>yaɾa</i>	‘man’	<i>ŋaɾu</i>	‘like a’
	<i>gugula</i>	‘platypus’	<i>baɖun</i>	‘proper’
	<i>bana</i>	‘water’	<i>gabun</i>	‘another’
	<i>waŋal</i>	‘boomerang’	<i>ɖaran</i>	‘two’

We can now build on this hypothesis to analyze some more data. One strategy that’s often good to use is to look for other words in which you already recognize a piece. Indeed it looks like examples (2g) and (2h) both have the piece *yaɾa* and a gloss that includes ‘man’. If we subtract this piece, we are left with two more bits we can now identify: *gabun* probably means ‘another’ and *ɖaran* ‘two’. This in turn suggests that we can identify the piece *baŋguyin* in (2i) as meaning ‘frog’.

At this point, we have a good idea how to analyze examples (2b–i), but we still haven’t cracked (2a). Example (2j) is still a problem as well, as so far, it doesn’t overlap with any of the other examples. But as soon as we go on to (2k), we start to get a clue, because there are two morphemes that we can now recognize in this word ‘boomerang’ and ‘two’, leaving only the final bit *bila* which therefore must mean ‘with’. We can now go back to (2j) and determine that *yugu* must mean ‘stick’. Example (2l) finally leads us back to example (2a): since we can identify a stretch in the middle of (2l) – the morpheme *gabun*, which we decided means ‘another’ – we can guess that *miɖa* means ‘camp’ and *ŋunu* means ‘from’. Why not the opposite, by the way? The reason is that so far it looks like the more semantically contentful morphemes, like ‘man’ and ‘frog’, always come first, and the less contentful come after; we might therefore hypothesize that morphemes like ‘from’ and ‘proper’ are suffixes in Dyirbal. And now we can finally go back to example (2a) and decide that the morpheme for ‘boy’ is *naɭga*. I leave it to you to analyze the last two examples, and check that our analysis so far is right.

I say ‘so far’ because we have only a tiny bit of data to work with here, and every morphological analysis is provisional on checking it against further data. Sometimes there are loose ends left after we’ve analyzed our data as much as we can. One loose end you might notice in our Dyirbal analysis is that it looks like there is no morpheme in any of our data that corresponds to a word like *a* in English. It’s impossible to know from this little data set whether Dyirbal has anything that corresponds to indefinite articles.

Challenge

Look at the following data from the language San Miguel Chimalpa Zoque, a Mixe-Zoquean language spoken in Mexico, and try your hand at breaking it down into individual morphemes. What do you think each morpheme means? (Johnson 2000: 413)

yədə	'this'
yəhə	'here'
yəhənəŋ	'towards here'
yəhəŋ	'over here, towards here'
yəhíŋ	'to this point, no farther'
yey	'soon, right now'
dedə	'that'
dehi	'there'
dehéŋ	'towards there'
dehíŋ	'from there'
dey	'now, then'

5.3 Affixes: Beyond Prefixes and Suffixes

Now that you've gotten your feet wet analyzing complex words in languages that might not be familiar to you, we can begin to look beyond prefixation and suffixation to morphological processes that we don't find in English. As we saw in Chapter 3, prefixes and suffixes are types of affixes that respectively go before or after a base, but these are not the only positions in which affixes can occur. This section will look at these different sorts of affixes.

5.3.1 Infixes

Infixes are affixes that are inserted right into a root or base. We saw an example in (1a) above. In Tagalog (Austronesian), it is possible to form intransitive verbs meaning 'become X' from adjectives by inserting the morpheme <um> after the first consonant of the adjective root. Example (4) shows how the words can be broken down, using the convention that the infix is enclosed in angled brackets (< and >):

(4)	g<um>anda	become
		beautiful
	h<um>irap	become
		difficult

Another example of infixation can be found in Karok. In Karok, a form of verb called the intensive is created by infixing the morpheme <eg> after the first consonant or cluster of consonants in the root, as in (5):

(5) *Karok (language isolate)* (Garrett 2001: 269)

Base verb		Intensive
la:y-	'to pass'	l<eg>a:y
ɬkyork ^w -	'to watch'	ɬky<eg>ork ^w
koʔmoy-	'to hear'	k<eg>oʔmoy
trahk-	'to fetch water'	tr<eg>ahk

Both examples of infixation that we've seen so far have had the infix right after the first consonant or consonant cluster of the base, but sometimes infixes can come near the end of the base as well. As the examples in (6) illustrate, in Hua, the negative infix <'a> comes before the last syllable of the verb root:

(6) *Hua (Trans-New Guinean)* (Haiman 1980: 195)

zgavo	'embrace'	zga<'a>vo	'not embrace'
harupo	'slip'	haru<'a>po	'not slip'
rvato-	'be nigh'	rva<'a>to	'not be nigh'

Hoava shows another interesting characteristic that sometimes appears with infixation. In Hoava an infix <in> is used to create nouns from adjectives or verbs. The <in> morpheme must be infixed if the base it attaches to begins with a consonant, but it appears as a prefix if the base begins with a vowel:

(7) *Hoava (Austronesian)* (Davis 2003: 39; Blevins 2014: 137)

a. to	'alive'	t<in>o	'life'
b. va-bobe	'fill'	v<in>abobe	'filled object'
c. ta-poni	'be given'	t<in>aponi	'gift'
d. asa	'grate'	<in>asa	'pudding of grated cassava'
e. edo	'happy'	<in>edo	'happiness'

Challenge

The following data are from Ulwa, a Misumalpan language spoken in Nicaragua. What is the morpheme that marks the possessed form and what do you notice about where it is placed in the word? (Davis and Tsujimura 2014: 203)

Base	Gloss	Possessed form
sana	'deer'	sanaka
bas	'hair'	baska
kii	'stone'	kiika
al	'man'	alka
siwanak	'root'	siwakanak
karasmak	'knee'	karaskamak
anaalaaka	'chin'	anaakalaaka

Infixation in English?

English doesn't have any productive processes of infixation, but there's one marginal process that comes close, which is affectionately

referred to by morphologists as ‘*fuckin*’ infixation’. In colloquial spoken English, we will often take our favorite taboo word or expletive – in American English *fuckin*g, *goddam*, or *frigging*, in British English *bloody* – and insert it into a base word:

abso-fuckin-lutely
fan-bloody-tastic
Ala-friggin’-bama

This kind of infixation is used to emphasize a word, to make it stronger.

What’s particularly interesting is that we can’t insert *fuckin*’ just anywhere in a word. Indeed, there are rather strict phonological restrictions on the insertion of expletives. Try inserting your favorite expletive into the following words:

Minnesota
elementary
onomatopoeia

Next, try it on the following words:

Texas
Georgia

Can you do ‘*fuckin*’ infixation’ with these words?

Now think up some other words and try ‘*fuckin*’ infixation’. Can you figure out what conditions the placement of the expletive?

5.3.2 Circumfixation and Parasynthesis

Another type of affix that occurs in languages is the **circumfix**. A circumfix consists of two parts – a prefix and a suffix that together create a new lexeme from a base. We don’t consider the prefix and suffix to be separate, because neither by itself creates that type of lexeme, or perhaps anything at all. This kind of affixation is a form of **parasynthesis**, a phenomenon in which a particular morphological category is signaled by the simultaneous presence of two morphemes.

One example of a circumfix can be found in Dutch, although Booij (2002: 119) says that it’s no longer productive. In Dutch, to form a collective noun from a count noun, the morpheme *ge-* is affixed before the base and *-te* after the base:

(8)	berg	‘mountain’	ge-berg-te	‘mountain chain’
	vogel	‘bird’	ge-vogel-te	‘flock of birds’

Neither *geberg* nor *bergte* alone forms a word – it’s only the presence of both parts that signals the collective meaning. Another example can be found in Tagalog, where adding *ka-* before and *-an* after a noun base *X* makes a noun meaning ‘group of *X*’:

(9) *Tagalog (Austronesian)* (Schachter and Otanes 1972: 101)

Intsik	'Chinese person'	ka-intsik-an	'the Chinese'
pulo	'island'	ka-pulu-an	'archipelago'
Tagalog	'Tagalog person'	ka-tagalog-an	'the Tagalogs'

Again, neither *ka* + noun, nor noun + *an*, has its own meaning in these words.

Two more examples of circumfixation can be found in (10), from Cavineña, and (11), from Kambera :

(10) *Cavineña (Tacanan)* (Guillaume 2008: 435–6; Bauer 2014: 127)

jutu	'to dress s.o.'	e-jutu-ki	'cloth'
sama	'to cure'	e-sama-ki	'medicine'
taru	'to stir'	e-taru-ki	'paddle'
teri	'to close sth'	e-teri-ki	'door'

(11) *Kambera (Austronesian)* (Klamer 1998: 245; Bauer 2014: 127)

ngùru	'murmur'	ka-ngùru-k	'to make a murmuring sound'
ndùru	'thunder'	ka-ndùru-k	'to make the sound of thunder'
mbàti	'drip'	ka-mbàti-k	'to make a dripping sound'
hètu	'sniff'	ka-hètu-k	'to make a sniffing sound'

Bauer (2014: 128) points out that Cavineña does have a prefix *e-* and a suffix *-ki*, but the meanings of these affixes when they are used alone are quite different from the meaning that we find in (10).

Challenge

Remember that in Chapter 3 we learned to draw word trees. Review Chapter 3, Section 2, and think about how we would have to draw word trees for words with circumfixes.

Circumfixes are only one sort of parasynthetic affix. Take a look at the data in (12) from Tzutujil and try to analyze them yourself. What are the two parts of the affix?

(12) *Tzutujil (Mayan)* (Dayley 1985: 176; Bauer 2014: 128)

tik-	'sow'	tijkoʔm	'sowing'
loq'	'buy'	lojq'oʔm	'item bought'

Here, it looks like we have an affix that consists of an infix and a suffix. Again, the critical thing that distinguishes parasynthesis from other sorts of affixation is that neither part of the affix can be used alone – at least with the relevant meaning.

5.3.3 Other Kinds of Affix

Occasionally in the literature on morphology we find reference to several other types of affix. For the most part, in this book we use different terms for these particular morphological processes, so here I will just mention the terms and refer you to the sections of this book where they are discussed:

- **interfixes:** These are what we have called linking elements. See Chapter 3, Section 3.4.2.
- **simulfixes:** This is another term for internal stem changes, which we will discuss in Section 5.4.
- **transfixes:** These are what we will call templatic morphology. See Section 5.6 below.

5.4 Internal Stem Change

Most of the forms of lexeme formation that we've looked at so far have involved adding something to a base or combining bases.² Some languages, however, have means of lexeme formation that involve changing the quality of an internal vowel or consonant of a base, root, or stem; sometimes this internal change occurs alone, and sometimes in conjunction with affixation of some sort. Such processes are called **internal stem change** or **apophony**.

5.4.1 Vowel Changes: Ablaut and Umlaut

Example (13) gives some words where internal vowels change:

- (13) a. *Manchu (Tungusic)* (Haenisch 1961: 34)
- | | | | |
|-------|----------|-------|----------|
| haha | 'man' | hehe | 'woman' |
| ama | 'father' | eme | 'mother' |
| amila | 'cock' | emile | 'hen' |
- b. *Muskogee (Muskogean)* (Haas 1940: 143)
- | | | |
|------|-------------|----------------|
| nis | 'to buy it' | Stem class I |
| ní:s | 'to buy it' | Stem class III |
| ni:s | 'to buy it' | Stem class IV |
- c. *German* (Lederer 1969: 25)
- | | | | |
|--------|-----------|------------|------------------|
| Bruder | 'brother' | Brüderlein | 'brother-dimin.' |
| Frau | 'woman' | Fräulein | 'woman-dimin.' |

In Manchu, in forming the female equivalent of a male noun, back vowels become front vowels. In Muskogee, verb stems have five forms each of which can be used in a number of verbal contexts (completive, incompletive, durative, and so on). Three of these forms are differentiated by the length and tonal patterns on their vowels. For class I, stems have short vowels and no special tonal accent. Class III stems have long vowels and falling tonal accents, and class IV have long vowels and no special accent. Morphological processes that affect the quality, quantity, or tonal patterns of vowels are often referred to as **ablaut**.

In German, when certain suffixes like the diminutive suffix *-lein* are added to a stem, the stem vowel becomes a front vowel. Historically, this fronting was a phonological process that occurred when a following suffix itself contained a front vowel; this process is called **umlaut**. Over time, the front vowels were lost in some suffixes or became back vowels, as is the case with the diminutive suffix *-lein* (pronounced [laɪn]).

A note on English Ablaut figures in a minor way in the morphology of English as well, as we can see in the past and past participle forms of verbs like sing (past sang, past pple. sung) or sit (past and past pple. sat). Since ablaut figures only in inflectional forms of English, though, we will postpone further discussion of it until Chapter 6.

2. The exception here is conversion, which we discussed in Chapter 3.

Another note on English

You might be interested to know that a historical process of umlaut is responsible for such singular/plural pairs in English as foot ~ feet or goose ~ geese. In earlier stages of English, the singular of the noun 'foot' was fot and its plural foti (similarly gos sg. ~ gosi pl.). The high front vowel of the plural suffix caused the vowel of the base to become front (so the singular and plural were then fot ~ feti and gos ~ gesi). The plural suffix was eventually lost, leaving the difference in vowels as the only signal of plurality in those words.

5.4.2 Consonant Mutation

In some languages morphological processes are signaled by changes in consonants rather than vowels in the base, root, or stem. Such processes are called **consonant mutations**. As with the vowel processes noted above, consonant mutations may occur alone or in conjunction with pre-fixes or suffixes.

- (14) a. *Seereer-Siin* (Niger-Congo) (McLaughlin 2000: 335)
- | | | | |
|--------|---------|---------------------|----------------|
| odon | 'mouth' | o ⁿ don | 'mouth-dimin.' |
| okawul | 'griot' | o ^g awul | 'griot-dimin.' |
| opacɗ | 'slave' | o ^m bacɗ | 'slave-dimin.' |
- b. *Chemehuevi* (Uto-Aztecan) (Press 1979: 21–2)
- | | | | |
|---------|-------|----------|-----------------|
| punikai | 'see' | navunika | 'see-reflexive' |
| tika | 'eat' | narika | 'eat-reflexive' |
| koa | 'cut' | naɣoa | 'cut-reflexive' |

In Seereer-Siin some noun diminutives are formed by replacing the first stop consonant in the stem, for example [d, k, p] in the words in (14a), with the corresponding prenasalized stop ([ⁿd, ^g, ^mb]). And in Chemehuevi, illustrated in (14b), the reflexive of a verb is formed by prefixing *na-* and changing the initial stop consonant of the root to a voiced continuant ([p] becomes [v], [t] becomes [r], and [k] becomes [ɣ]).

Challenge

The following data come from a language isolate called Nivkh, which is spoken in Siberia (Spencer 1991: 19; Davis and Tsujimura 2014: 210):

Transitive	Intransitive	Gloss
ɾʌŋzʌɬɔd	tʌŋzʌɬɔd	'weigh'
ɣavud	q ^h avud	'warm up'
ɣesqod	kesqod	'burn something/oneself'
vakzd	pakzd	'lose/get lost'
rad	t ^h ad	'bake'

Use the IPA chart at the beginning of the book to help you figure out how the consonant mutation that relates the transitive and intransitive forms works.

5.5 Reduplication

Reduplication is a morphological process in which all or part of a word (typically the base, but not always) is repeated. Some examples are given in (15):

- (15) a. *Hausa* (Afro-Asiatic) (Newman 2000: 42)
- | | | | |
|------|-----------|-----------|-----------------|
| bāya | 'behind' | bāya bāya | 'a bit behind' |
| gàba | 'forward' | gàba gàba | 'a bit forward' |
| ƙasà | 'below' | ƙasà ƙasà | 'a bit below' |

b. *Samoan (Austronesian)* (Mosel and Hovdhaugen 1992: 229)

'apa	'beat, lash'	'apa'apa	'wing, fin'
au	'flow on, roll on'	auau	'current'
solo	'wipe, dry'	solosolo	'handkerchief'

(15a) and (15b) illustrate **full reduplication**, a process by which an entire base is repeated. In the case of Hausa, full reduplication is used to form what's called an **attenuative**, which is a form meaning 'sort of' or 'a little bit'. In Samoan full reduplication is used to form nouns from verbs.

Some languages have what is called **partial reduplication** in which only part of the base is repeated. The example in (16) comes from Samoan.

(16) *Samoan* (Mosel and Hovdhaugen 1992: 223)

lafo	'plot of land'	lalafo	'clear land'
lago	'pillow, bolster'	lalago	'rest, keep steady'
pine	'pin, peg'	pipine	'secure with pegs'

In (16) you can see that partial reduplication in Samoan repeats the first consonant and vowel of the base; this process derives verbs from nouns. Mokilese also has partial reduplication, but in this case what is reduplicated is a heavy syllable, that is, a syllable that either has a long vowel or ends in a consonant:³

(17) *Mokilese (Austronesian)* (Harrison 1973; Inkelas 2014: 170)

pɔdɔk	'plant'	pɔd-pɔdɔk	'planting'
soorɔk	'tear'	soo-soorɔk	'tearing'
diar	'find'	dii-diar	'finding'
andip	'spit'	and-andip	'spitting'

Partial reduplication need not repeat the initial part of a base; it may also in some languages repeat the final part of the base, as the example from Teton Dakota in (18) illustrates:

(18) *Teton Dakota (Siouan)* (Shaw 1980: 321)

wa+ksà	'cut with sawing motion'	wa+ksà-ksà	'slice up'
wač ^h í	'dance'	wač ^h íč ^h í	'jump up and down'

In this dialect of Dakota, the final syllable of the verb root can be reduplicated to indicate iterative or repetitive action.

Another sort of process that can be considered a form of reduplication is one where a base is fully or partially repeated, but with the substitution of one or more segments, something which Inkelas (2014: 171) refers to as 'echo' reduplication. One dialect of English actually has a process that illustrates this. In Yiddish-English you can replace the beginning of a word with the sequence [ʃm] to express irony or derision, as illustrated in (19):

(19) book schmook, fix schmix, Jennifer schmennifer

Reduplication is an interesting process in terms of structure, as we have seen above, but it is also interesting semantically in that it sometimes expresses meanings that can be considered iconic. By 'iconic', we

3. We will look more at the interaction of reduplication and syllable structure in Chapter 9.

mean cases in which the form of the derived word suggests its meaning. When reduplication is semantically iconic, it can express plurality, intensity, augmentation, repetition, or other meanings suggesting 'more-ness'. It is important to point out, though, that the meaning conveyed by reduplication need not be iconic. Cross-linguistically, processes of reduplication can be found which express exactly the opposite of 'more-ness', diminutives, for example, as well as meanings that have nothing to do with size or quantity. Indeed, as the examples in (15b) and (16) illustrate, reduplication can be used for canonical derivational purposes such as forming nouns from verbs or verbs from nouns.

Challenge

Does English have any other sorts of reduplication? Consider the following examples from English:

willy-nilly
hocus-pocus
mumbo-jumbo
hanky-panky
hodge-podge
handy-dandy
hoity-toity
helter-skelter

Can you think of more examples of this sort? Do you think that words like these express a process of reduplication in English? If so, is it productive? If not, why not?

5.6 Templatic Morphology

Consider the data in (20):

- (20) *Arabic (Afro-Asiatic)* (McCarthy 1979: 244; 1981: 374)
- | | |
|--------|------------------------------------|
| katab | 'wrote' |
| kattab | 'caused to write' |
| kaatab | 'corresponded' |
| ktatab | 'wrote, copied' |
| kutib | 'was written' (perfective passive) |

All of the words in (20) have something to do with writing, and all share the consonants *ktb*, although in a couple of the forms, there's more than one *t*. All the active verb forms have the vowel *a*; the passive verb form has the vowels *u-i*. Each word has a different pattern of vowels and consonants, and each expresses a slightly different concept. What we find in Arabic is called **templatic morphology** or **root and pattern morphology**.

In Arabic, the root of a word typically consists of three consonants (like *ktb*), the **triliteral root**, which supply the core meaning. These three consonants may be interspersed with vowels in a number of different ways to modify the meaning of the root. The precise pattern of consonants (abbreviated C) and vowels (abbreviated V) – sometimes called the **template** – can be associated with specific meanings. For example, the pattern CVCVC is the non-causative non-reciprocal form of the verb, but the pattern CVCCVC conveys a causative meaning, and the pattern CVVCVC a reciprocal meaning (we can take *correspond* to mean something like ‘write to each other’). Each of these template patterns is called a **binyan**, a term which comes from traditional Hebrew grammar. The specific vowels that get interspersed between the consonants in these patterns can contribute inflectional meanings; so the vowel *a* is used in active forms, and the vowels *u-i* in passive forms. Roots in Arabic are occasionally called **transfixes** because some morphologists look at them as affixes that occur discontinuously across the word.

Root and pattern morphology is very characteristic of the Semitic family of languages, which includes Arabic and Hebrew. But it can also be found in other languages, for example Cupeño. Cupeño verbs can have a form called the **habilitative**, which means something like ‘can V’:

(21) *Cupeño (Uto-Aztecan)* (McCarthy 1984: 309)

a.	čál	‘husk’	čáʔaʔal	‘can husk’
	tééw	‘see’	téʔeʔew	‘can see’
	həlyəp	‘hiccup’	həlyəʔəʔəp	‘can hiccup’
	kəlaw	‘gather wood’	kəláʔaʔaw	‘can gather wood’
b.	páčik	‘leach acorns’	páciʔik	‘can leach acorns’
	čáspəl	‘mend’	čáspəʔəl	‘can mend’

One way of looking at the habilitative forms in Cupeño is that they conform to templates like those in (22), where the vowel that is bolded is the stressed vowel:

(22) Template for Cupeño habilitatives

- a. (CV)CV**ʔ**V**ʔ**VC
- b. CVC(C)V**ʔ**VC

If the only or the second vowel is the stressed vowel, as is the case in the examples in (21a), the habilitative form adds two more syllables, each of which start with [ʔ]. The final vowel of the stem is spread to the new syllables. The situation is slightly different if the first vowel of two is the stressed vowel, as the examples in (21b) show. In that case, the template has only one more syllable than the stem, again with the glottal stop as the consonant, and the vowel supplied by the last stem vowel.

A final example of templatic morphology comes from another Native American language, Sierra Miwok. In this language, new words can be derived by adding a suffix which then supplies a specific template for the base. Consider the examples in (23):

- (23) *Sierra Miwok (Miwok-Costanoan)* (Smith 1985: 365, 371–2)
- | | | | | |
|----|--------|-----------------------|---------------|----------------------|
| a. | petja | ‘drop several things’ | peeťaj-tee-ny | ‘string out’ |
| | halki | ‘hunt’ | haalik-tee-ny | ‘hunt along a trail’ |
| b. | hulaw | ‘forget’ | hulwaw-we | ‘be late’ |
| | ?okiih | ‘beg for food’ | ?okhih-he | ‘be pitiful’ |
| c. | hywaat | ‘run’ | hywattatt | ‘run around’ |
| | hyleet | ‘fly, be in the air’ | hylettett | ‘flop about (fish)’ |

The examples in (23a) show that the suffix *-tee-ny*, which forms what Smith calls the ‘linear distributive’, makes the verb stem conform to a template of the form CVVCVC. The forms in (23b) have a suffix with the form Ce, where the C is the last consonant of the verb stem. In addition, the verb stem is made to conform to the pattern CVCCVC. Finally, in (23c) verb stems are made into derived forms that mean something like ‘X around’ just by making them conform to a template that looks like CVCVCCVCC, with no suffix added.

Challenge

Consider the following data from modern Hebrew (Afro-Asiatic) (Davis and Tsujimura 2014: 195):

lamad	‘learn’	limed	‘teach’
takan	‘be straight’	tiken	‘repair’
yavan	‘Greece’	yiven	‘Hellenize’
targum	‘translation’	tirgem	‘translate’

Describe the process that is used to derive the verbs in the second column from their bases in the first column.

5.7 Subtractive Processes

Generally we think of lexeme formation as an additive process. We start out with a base and add affixes (whether prefixes, suffixes, infixes, or circumfixes) or compound two or more bases or reduplicate part of the base, resulting in a derived word that is longer than the original base. Not all lexeme formation is additive, however. There are some languages that have processes that we might call subtractive in the sense that the derived lexeme consists of a base from which something has been removed. One example can be found in Koasati:

- (24) *Koasati (Muskogean)* (Martin 1988: 230–1)
- | | | |
|-----------------|---------------|--------------------------|
| <i>singular</i> | <i>plural</i> | |
| pitáf-li-n | pít-li-n | ‘to slice up the middle’ |
| tiwáp-li-n | tíw-wi-n | ‘to open something’ |
| icoktaká:-li-n | icokták-li-n | ‘to open one’s mouth’ |
| misíp-li-n | mís-li-n | ‘to wink’ |

The plural is formed by subtracting the final vowel and consonant of the base.

Challenge

Why do we say that the process in (24) is subtractive? Consider an alternative hypothesis in which the singular is formed from the plural bases (*tɪw*, *pɪt*, etc.) by adding a morpheme. Why would this not be a good hypothesis?

Unusual as **subtractive morphology** may seem to speakers of English, we actually have something comparable. It's common in English to form nicknames (**hypocoristics**, if you want the technical term!) by subtracting part of the full name. Consider the names in (25) with their hypocoristics:

(25)	Patrick	Pat
	Barbara	Barb
	Mitchell	Mitch
	Susan	Sue
	Alan	Al
	Benjamin	Ben

To form a nickname in English, we typically subtract all but the first CV or CVC sequence of the name. Sometimes we add the suffix *-y* or *-ie* as well, as in *Patty* from *Patricia* or *Allie* from *Allison*. So we can think of nickname formation as a sort of subtractive process as well, with or without accompanying affixation.

Summary

In this chapter we have completed our survey of the different types of rules that can be used in forming new lexemes in the languages of the world. We have gone beyond prefixation, suffixation, compounding, and conversion to add new types of affixes (infixes, circumfixes), and new processes like internal stem change (ablaut, umlaut, and consonant mutation), reduplication (full and partial), templatic and subtractive morphology.

Exercises

1. The Austronesian language Leti has a process that derives nouns meaning 'the act of V-ing' from verbs. Consider the data below (from Blevins 1999: 390):

kakri	'cry'	kniakri	'the act of crying'
pali	'float'	pniali	'the act of floating'
sai	'climb'	sniai	'the act of climbing'
teti	'chop'	tnieti	'the act of chopping'
vaka	'ask (for)'	vniaka	'the act of asking'
va-nunsu	'knead'	vnianunsu	'massage' = 'the act of kneading'

- a. What is the morphological rule that creates nouns from verbs in Leti? What kind of a rule is it?
- b. Now consider the following forms:
- | | | | |
|------|---------|--------|------------------------------------|
| atu | 'know' | niatu | 'knowledge' = 'the act of knowing' |
| odi | 'carry' | niodi | 'the act of carrying' |
| osri | 'hunt' | niosri | 'hunt' = 'the act of hunting' |

Divide these new words into morphemes and discuss what changes you need to make to the morphological rule you wrote for part (a) in order to account for this new data.

2. The examples below are from Yurok (Algic) (data from [Garrett 2001](#): 274):

kep'eł	'housepit'	kep'kep'eł	'there are several housepits'
ket'ul	'there's a lake'	ket'ket'ul	'there's a series of lakes'
pegon	'to split'	pegpegon	'to split in several places'
siton	'to crack' (intrans.)	sitsiton	'to crack several times'
tekun	'to be stuck together'	tektekun	'to be stuck together in several places'

Write a word formation rule that derives the Yurok forms in the second column from the corresponding base in the first column. Make sure to include both the structural and semantic effects of the rule. What kind of a morphological rule is this?

3. Consider the data below from Kannada (Dravidian) (data from [Sridhar 1990](#): 268):

a:Ta	'game'	a:Ta-gi:Ta	'games and the like'
huli	'tiger'	huli-gili	'tigers and the like'
sphu:rti	'inspiration'	sphu:rti-gi:rti	'inspiration, etc.'
autaNa	'banquet'	autaNa-gi:taNa	'banquet, etc.'

Try to write a morphological rule that derives the words in the second column from the bases in the first column. What kind of morphological rule is this? How does it differ from other morphological rules we've looked at in this chapter?

4. In Gta? (Austro-Asiatic) a number of different forms can be derived from a noun base, as the examples here show (data from [McCarthy 1983](#)):

kitoŋ	'god'
kataŋ	'being with powers equal to "kitoŋ"'
kitiŋ	'being smaller, weaker than "kitoŋ"'
kutaŋ	'being other than "kitoŋ" (e.g., spirits, ghosts)'
kesu	'wrapper worn against cold'
kasa	'cloth equivalent to "kesu" in size and texture'
kisi	'small or thin piece of cloth'
kusa	'any other material useable against cold'

Propose an analysis of this process. What kind of word formation rule is at work here?

5. The following words, taken from Yu (2004: 620), are characteristic of the speech of the character Homer Simpson from the animated TV show *The Simpsons*. In this data, Homer Simpson seems to display a process of infixation:

saxomaphone	'saxophone'	Missimassippi	'Mississippi'
telemaphone	'telephone'	Alamabama	'Alabama'
wondermaful	'wonderful'	diamalectic	'dialectic'
feudamalism	'feudalism'	Michamalangelo	'Michaelangelo'
secrematery	'secretary'	territatory	'territory'

Is this like real cases of infixation that we saw in this chapter? If so, why? If not, why not? Try to formulate a precise rule for Homer Simpson infixation.

6. The following examples from Amharic (Afro-Asiatic) illustrate a form of language disguise or play language (like Pig Latin) used by young women in the Ethiopian capital, Addis Ababa (McCarthy 1984: 306):

g ^w aro	'backyard'	g ^w ayrər
gin	'but'	gaynən
mətt'a	'come'	mayt'ət
kifu	'cruel'	kayfəf
həd	'go'	haydəd
man	'who'	maynən

Figure out the morphological rule that creates the play language version (the third column) of the Amharic words. What kind of morphological rule is this?

7. Consider the following, from Alabama (Muskogean) (Hardy and Montler 1988: 394):

salatli	'slide once'	salaali	'slide repeatedly'
haatanatli	'turn around once'	haatanaali	'turn around repeatedly'
noktilifka	'choke once'	noktiliika	'choke repeatedly'

Describe the word formation process that derives the words in the second column from those in the first column. What kind of morphological process is this?

8. The following data come from Koasati (Muskogean) (Kimball 1991: 351). Write a word formation rule for the process that they illustrate:

molápkán	'to gleam'	molalápkán	'to flash'
bolótín	'to shake'	bololótín	'to shake with fear'
wacíplín	'to feel a stabbing pain'	wacícíplín	'to feel repeated stabbing pains'
konótín	'to roll'	kononótín	'to quiver fatly'
watóhlin	'to clabber'	watotóhlin	'to jiggle like jello'

9. The language Semai (Austro-Asiatic) derives what Diffloth (1976) calls expressive forms from nouns using a process of reduplication. Consider the data below and try to formulate the rule of reduplication (data from Diffloth 1976: 252–6):

dŋəh	dhdŋəh	'appearance of nodding constantly'
dyɔːl	dldyɔːl	'appearance of object which goes on floating down'
taʔəh	thtaʔəh	'appearance of a large stomach constantly bulging out'
ghə:p	gpghə:p	'irritation on skin'

10. Examine the data below from Huastec (Uto-Aztecan) (data from [Edmonson 1995](#): 380–2). Identify the morphemes and try to assign them meanings. Discuss anything that strikes you as unusual in these data.

a. ʔu ɕiʔikme:l	'it becomes sweet'
b. ʔu ɕiʔikme:	'it became sweet'
c. ʔu ɕiʔikme:nek	'it has become sweet'
d. ʔu ɕiʔikme:nekiɕ	'it has already become sweet'
e. ʔin ɕiʔikme:θa:l	'he makes it become sweet'
f. ʔin ɕiʔikme:θaʔ	'he made it become sweet'
g. ʔin ɕiʔikme:θa:mal	'he has made it become sweet'
h. ʔan ɕiʔikme:θom	'the one who sweetens'
i. ɕiʔikme:θa:mehiɕ	'it has already been made sweet'
j. ʔin ɕiʔikme:θanči	'he made it become sweet for someone'
k. ʔin ɕiʔikme:θanča:malɕ	'he has already made it become sweet for someone'

11. The following data are from the Hualapai (Yuman) (from [Watahomigie, Bender, and Yamamoto 1982](#): 197–8). Discuss the process that is used to form the adjectives/verbs from nouns.

chud	'winter'	chu:dk	'wintery'
guwí	'cloud'	guwi:k	'cloudy'
i'i	'wood'	i'i:k	'woody'
màɕakwíd	'whirlwind'	màɕakwi:dk	'whirlwindy'
gínya	'younger sibling'	gi:nyk	'to have a younger sibling'
kwàsivɕí	'fence'	kwàsivdi:k	'to fence'

6 Inflection

CHAPTER OUTLINE

KEY TERMS

person
number
gender
case
tense
aspect
inherent
contextual
paradigm
evidentiality
mirativity

In this chapter you will learn about inflection, the sort of morphology that expresses grammatical distinctions.

- ◆ We will look at a wide variety of types of inflection, including number, person, gender and noun class, case, tense and aspect, voice, mood and modality.
- ◆ We will learn what morphologists mean by a 'paradigm' and look at patterns within paradigms.
- ◆ And we will consider whether it is always clear where to draw the line between inflection and derivation.

6.1 Introduction

At the outset of this book we divided morphology into two domains: **inflectional** and **derivational** word formation. In the last three chapters, we have concentrated on derivational word formation – types of word formation that create new lexemes. In this chapter, we turn our attention to inflectional word formation.

Inflection refers to word formation that does not change category and does not create new lexemes, but rather changes the form of lexemes so that they fit into different grammatical contexts. As we'll see in detail below, grammatical meaning can include information about number (singular vs. plural), person (first, second, third), tense (past, present, future), and other distinctions as well. In this chapter, we will first survey different forms of inflection that can be found both in English and in the languages of the world, and then look at the ways in which inflection can work.

A word before we start though. We've seen that new lexemes can be derived using all sorts of different formal processes of word formation: affixation, compounding, conversion, internal stem change, reduplication, templatic morphology, and so on. Inflectional word formation makes use of almost all of these types of word formation rules as well, with the possible exceptions of compounding and subtractive processes. That is, just as languages may have derivational affixes that form new lexemes, they may have inflectional affixes that make those lexemes suited for one grammatical context or another; similarly, languages may have rules of reduplication for either derivational purposes or inflectional purposes. In other words, we might say that *form* (the type of rule or process) is independent of *function* (derivation or inflection). Keep this in mind; many of the examples that I'll use to illustrate points below make use of affixation, but in many cases I could have chosen examples with reduplication or internal stem change as well.

6.2 Types of Inflection

Native speakers of English are often surprised at the kinds of inflection that can be found in languages – English is a language that has relatively little inflection, as languages go. So we'll start by surveying some of the types of inflection that can be found in the languages of the world.

6.2.1 Number

Perhaps the most familiar inflectional category for speakers of English is **number**. In English, nouns can be marked as singular or plural:

- | | | |
|-----|-----------------|----------------------------|
| (1) | <i>Singular</i> | cat, mouse, ox, child |
| | <i>Plural</i> | cats, mice, oxen, children |

Although the vast majority of nouns pluralize in English by adding -s (or in terms of sounds, one of the variants [s], [z], or [əz]), some nouns form their plurals irregularly. We will return to the issue of regular *versus*

irregular inflections shortly. In English, it is required to mark the plural on nouns in a context in which more than one of that noun is being discussed (*I have six beagles*). This is not the case in all languages, as our Mandarin Chinese example in Chapter 1 illustrated.

Some languages distinguish a third category of number in addition to singular and plural. For example, in Yup'ik, nouns inflect not only for singular and plural, but also for what is called **dual**. This is a number marking that means 'two':

- (2) *Yup'ik (Eskimo-Aleut)* (Mithun 1999: 79)
- | | | | |
|-------|------------------------|----------|-------------------------|
| qayaq | 'kayak' | paluqtaq | 'beaver' |
| qayak | 'two kayaks' | paluqtak | 'two beavers' |
| qayat | 'three or more kayaks' | paluqtat | 'three or more beavers' |

As we'll see soon, some languages can make the singular/dual/plural distinction on verbs, as well as on nouns.

6.2.2 Person

Students of Indo-European languages like Latin or German know that verbs in those languages are marked for the inflectional category of **person**: that is, verbs exhibit different endings depending on whether the subject of the sentence is the speaker (first person), the hearer (second person), or someone else (third person); frequently number is also expressed as well as person:

- (3) a. *Latin: amāre* 'to love'
- | | | | | | |
|-----------------|-----|-----------|---------------|-----|---------------|
| <i>Singular</i> | 1st | amō (-o) | <i>Plural</i> | 1st | amāmus (-mus) |
| | 2nd | amās (-s) | | 2nd | amātis (-tis) |
| | 3rd | amat (-t) | | 3rd | amant (-nt) |
- b. *German: sagen* 'to say'
- | | | | | | |
|-----------------|-----|-------------|---------------|-----|-------------|
| <i>Singular</i> | 1st | sage (-e) | <i>Plural</i> | 1st | sagen (-en) |
| | 2nd | sagst (-st) | | 2nd | sagt (-t) |
| | 3rd | sagt (-t) | | 3rd | sagen (-en) |

Speakers of Indo-European languages may, however, be less familiar with marking person on nouns. It is not unusual for languages to mark person on nouns to show possession, something we do in English with separate possessive pronouns. For example, Mohawk uses prefixes to mark person on nouns:

- (4) *Mohawk nouns (Iroquoian)* (Mithun 1999: 69)
- | | | | |
|-----------------|------------|-----------------|--------------------|
| <i>Singular</i> | 1st person | khnia'sà:ke | 'my throat' |
| | 2nd person | shnia'sà:ke | 'your throat' |
| | 3rd person | iehnia'sà:ke | 'her throat' |
| | | rahnia'sà:ke | 'his throat' |
| <i>Plural</i> | 1st person | iakwahnia'sà:ke | 'our throats' |
| | 2nd person | sewahnia'sà:ke | 'your pl. throats' |
| | 3rd person | kontihnia'sà:ke | 'their F. throats' |
| | | ratihnia'sà:ke | 'their M. throats' |

Mohawk and other languages also show another kind of person marking that we don't have in English, a distinction between the inclusive and exclusive forms of the first person plural. In an **inclusive** form only the speaker and the hearer are included. The **exclusive** form includes the speaker and others, but not the hearer. So the inclusive form of the first person could be thought of as a combination of first and second person marking, and the exclusive as a combination of first and third person marking. This distinction can be marked in Mohawk as well, specifically with prefixes on verbs:

- (5) *Mohawk verbs* (Mithun 1999: 70)
- | | | |
|-----------------------------|---------------|--------------------------------------|
| <i>1st person inclusive</i> | tewahià:tons | 'we all (you pl. and I) are writing' |
| <i>1st person exclusive</i> | iakwahià:tons | 'we all (they and I) are writing' |

As we mentioned above, it is also possible to mark verbs if the subject consists of exactly two people. So in addition to the inclusive and exclusive forms in (5), Mohawk also has first person dual inclusive and exclusive forms (Mithun 1999: 70):

- (6) *Dual inclusive* tenihià:tons 'we two (you and I) are writing'
- Dual exclusive* iakenihià:tons 'we two (s/he and I) are writing'

English verb forms are ambiguous with respect to these distinctions. If I say "we write," neither the form of the verb nor the form of the pronoun makes explicit whether the hearer is included or not, or how many other than the speaker are involved (although of course we can make the distinction in a roundabout way, if we need to!).

6.2.3 Gender and Noun Class

If you've studied French, Spanish, German, Latin, Russian, or another Indo-European language, you're probably familiar with the concept of **gender**. In languages that have **grammatical** gender, nouns are divided into two or more classes with which other elements in a sentence – for example, articles and adjectives – must agree. We use French and German as our examples here:

- (7) a. *French*
- | | | | |
|------------------|--------|-----------------|---------|
| <i>Masculine</i> | | <i>Feminine</i> | |
| homme | 'man' | femme | 'woman' |
| rat | 'rat' | souris | 'mouse' |
| bureau | 'desk' | table | 'table' |
- b. *German*
- | | | | | | |
|------------------|---------|---------------|---------|-----------------|---------|
| <i>Masculine</i> | | <i>Neuter</i> | | <i>Feminine</i> | |
| Mann | 'man' | Kind | 'child' | Frau | 'woman' |
| Hund | 'dog' | Pferd | 'horse' | Maus | 'mouse' |
| Tisch | 'table' | Pult | 'desk' | Mauer | 'wall' |

French has two genders, masculine and feminine. German has three genders, masculine, feminine, and neuter. While sometimes the real-world gender of the noun's referent determines the grammatical gender of the noun – that is, the class that the noun belongs to – in many more cases nouns are assigned to genders with some degree of arbitrariness. So while the words for 'man' and 'woman' in both languages are masculine and feminine respectively, in accordance with **natural** gender, the

assignment of various animal names to gender classes is quite arbitrary. Rats, mice, dogs, and horses of course have natural gender – in the real world they are either male or female – but French and German grammar places them in a gender independent of their natural genders. In French, rats are masculine, but mice feminine. In German, dogs are masculine, horses neuter, and mice feminine. Inanimate nouns have no natural gender, but they are nevertheless classed as either masculine or feminine in French, and as any of the three genders in German.

Assignment to gender classes sometimes seems completely arbitrary, but it is not always as arbitrary as you might think. For example, in German all nouns derived with the derivational suffixes *-ung*, *-keit*, *-heit*, and *-schaft* are feminine, regardless of the gender of their bases. All of these suffixes form abstract nouns, so it's possible to say that derived abstract nouns are always feminine. Similarly, the diminutive suffixes *-chen* and *-lein* produce neuter nouns, regardless of the base they attach to. In other languages, nouns may be assigned to a gender based on their phonological shape. For example, Hausa (Afro-Asiatic) has masculine and feminine genders. Nouns for male humans and large animals are masculine and those for female humans and large animals are feminine in accordance with natural gender. The rest of the nouns are assigned arbitrarily to one of the genders, but Corbett (1991: 53) notes that there is a strong tendency for nouns that end in the vowels *-aa* to be feminine.

For many nouns in French and German, neither the meaning of the noun nor its phonological form signals its gender. In other words, there are no suffixes or other marks right on the nouns to tell us their genders (life would be much easier for second language learners of these languages if there were!). Rather, we can tell what the gender of the noun is by other elements in a sentence that are in **agreement** with a noun. So in French, the definite article *le* is used with masculine nouns (*le bureau*), and *la* with feminines (*la table*). Similarly, in German the definite article *der* is used with masculines (*der Tisch*), *das* with neuters (*das Pferd*), and *die* with feminines (*die Maus*).

The gender systems we are most familiar with are typical of Indo-European languages, and occur in other language families as well, but there are many languages outside of Indo-European that exhibit genders or **noun classes** based on distinctions other than (or in addition to) masculine, feminine, and neuter. Languages may have human and nonhuman classes or classes for rational beings as opposed to everything else. In the Algic family of languages, noun classes are based not on masculine and feminine but on animacy. Words for people belong to the animate class, but so, for example, do spirits and animals. And while most words for inanimate things belong to the inanimate class, some belong to the animate class; for example, the nouns for 'snowshoe' and 'button' in the Algic language Ojibwa¹ belong to the animate noun class (Corbett 1991: 20). In other words, while there is a

1. Corbett (1991) refers to this language as Ojibwa. Ethnologue classes Ojibwa as a 'macrolanguage'. Individual dialects are frequently known these days by 'autonyms', that is, the name that the speakers themselves use for the language. Nishnaabemwin, a language that we will discuss in Chapter 7, is one dialect of Ojibwa.

partial semantic basis for the two classes, assignment to the animate and inanimate classes can still be arbitrary.

Languages are not limited to two or three classes. Consider Yuchi:

(8) *Yuchi (language isolate)* (Mithun 1999: 103)

Noun class	Gender of speaker	
1.	M	singular or plural Yuchi, except certain female relatives
2.	M	singular female Yuchi relative, same or descending generation (sister, daughter, niece, granddaughter)
	F	any female Yuchi of same or descending generation
3.	F	singular male Yuchi relative, same or descending generation (brother, son, nephew, grandson)
4.	M, F	singular female Yuchi relative, ascending generation (mother, aunt, grandmother)
5.	F	singular male unrelated Yuchi, or plural Yuchis of same or descending generation
6.	F	singular male Yuchi of ascending generation (father, uncle, grandfather, husband, or as a term of respect for Yuchis of ascending generation)
7.	M, F	any non-Yuchi(s), animals
8.	M, F	vertical inanimate objects
9.	M, F	horizontal inanimate objects
10.	M, F	round inanimate objects

In Yuchi, nouns fall into classes based on whether they denote humans, animals, or inanimate objects with certain shapes. For example, Class 1 contains nouns used by men to denote Yuchi people, except close female relatives. The gender of a noun is indicated by a number of things, one of which is the article suffix used with the noun. For example, *gɔn'te-nɔ́* means 'the Yuchi man' and *gɔn'te-wənɔ́* 'the non-Yuchi man'. 'The tree' is *yáfa* but *ya-ʔé* is 'the log'. Class 7 contains nouns used by either men or women to denote animals or people who are not Yuchi. Other classes distinguish Yuchis from other humans, and Yuchis related to the speaker from those unrelated to the speaker.

Related to gender or noun class systems are what some linguists call **classifiers**, which are words or morphemes that are used in noun phrases with quantifiers (like *all*, *some*), number words, demonstratives (*this*, *that*) and the like. The classifier for a particular noun is chosen according to various semantic criteria that might include animacy, shape (e.g., long, flat), consistency (e.g., wet, flexible), or kind (e.g., fruit, book) (Aikhenvald 2000: 93; Velupillai 2012: 172). Unlike gender or noun class, classifiers typically don't agree with nouns or determiners. They just serve to categorize nouns in noun phrases. As we saw in Chapter 1 (examples (6), (7)), Mandarin uses a system of classifiers in noun phrases, so that if we want to talk about some number of giraffes, we need to say 'one *zhi* giraffe', where *zhi* is the classifier for animals of this sort. However, if we were counting books, we would need to use the classifier *ben* between the numeral and the noun *book*.

6.2.4 Case

Case is another grammatical category that may affect nouns (or whole noun phrases). In languages that employ the inflectional category of case, nouns are distinguished on the basis of how they are deployed in sentences, for example, whether they function as subject, direct object, indirect object, as a location, time, or instrument, or as the object of a preposition. In Latin, for example, nouns must be inflected in one of five cases, with singular and plural forms for each case:

(9) *Latin:*

<i>Singular</i>	<i>stella</i> ‘star’ (F)	<i>puer</i> ‘boy’ (M)
Nominative	stella	puer
Genitive	stellae	puerī
Dative	stellae	puerō
Accusative	stellam	puerum
Ablative	stellā	puerō
<i>Plural</i>		
Nominative	stellae	puerī
Genitive	stellārum	puerōrum
Dative	stellīs	puerīs
Accusative	stellās	puerōs
Ablative	stellīs	puerīs

The **nominative** case forms are used for the subject of the sentence. **Accusative** is generally used for the direct object and **dative** for the indirect object. **Genitive** is used for the possessor (e.g., *the boy’s shirt*). **Ablative** is used for the objects of prepositions (e.g., *cum* ‘with’, *dē* ‘from’), although some prepositions take objects in the accusative case (*ad* ‘to’, *post* ‘after’). Latin displays what is commonly called a **nominative–accusative case system**.

In this sort of system, subjects of verbs are nominative, whether the verbs are transitive (i.e., they take an object) or intransitive (they don’t take an object):

- (10) a. Puer amat puellam
 boy.NOM loves girl.ACC
 ‘The boy loves the girl’
 b. Puer it
 boy.NOM goes
 ‘The boy goes’

Less frequent is a kind of case marking system called an **ergative–absolutive** system. In this kind of system, the subject of a transitive verb gets a case called the **ergative**. The subject of an intransitive verb gets a case called the **absolutive**, which is also the case used for the direct object of a transitive verb. The examples in (11) from Georgian (Kartvelian) illustrate an ergative–absolutive case marking system (Whaley 1997: 163):²

2. Interestingly, Georgian has an ergative–absolutive case marking system in the perfect tense, but in the present tense it has a nominative–accusative system.

- (11) a. Student-i mivida
 student-ABS went
 ‘The student went’
- b. Student-ma ceril-i dacera
 student-ERG letter-ABS wrote
 ‘The student wrote the letter’

You can see the two systems compared schematically in (12):

(12)

	Nominative–Accusative	Ergative–Absolutive
Subject of transitive verb	Nominative	Ergative
Subject of intransitive verb	Nominative	Absolutive
Object of transitive verb	Accusative	Absolutive

Ergative–absolutive case systems are less frequent in the languages of the world than nominative–accusative systems, but they do occur in the Pama-Nyungan languages of Australia (e.g., Dyirbal), in the Tsimshianic languages of North America (e.g., Sm’algayax, spoken in British Columbia), in the language isolate Basque, as well as in Kartvelian languages like Georgian.

The Leipzig Glossing Rules

As you’ve already seen, morphologists frequently study word formation phenomena from languages that we don’t speak ourselves and may not have studied directly. To do so, we rely on grammars and articles written by experts. Data in such sources are most helpful and informative when they are **glossed**. By **glossing**, we refer to a morpheme-by-morpheme translation, frequently written right under an example in the language in question. The gloss is often then followed by a colloquial translation, as in the example you saw in (11). Until recent years there was no agreed upon convention for what abbreviations to use for particular inflectional or derivational meanings, or what symbols to use to separate morphemes in the language examples and meanings in the glosses. The Leipzig glossing rules are an attempt to rectify this issue. These days, linguists are encouraged to follow the guidelines for glossing that can be found at www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf. I won’t give you the entire list of rules, but will mention just a few that are useful for the student morphologist.

- Morphemes and their glosses are separated by hyphens and aligned so that the gloss appears just below the morpheme it translates.
- A colloquial translation occurs below the gloss, enclosed in single quotes.
- Clitics (bound morphemes that are more loosely attached to their bases than suffixes are – see **Chapter 8**), as opposed to affixes, are separated from their host by an = sign.

- When the gloss has more parts than the morpheme it is translating, the meanings are separated by a period, as in the Turkish and Latin examples below.
- Grammatical morphemes such as inflections are written in small capitals.

Turkish *çık-mak*

come.out-INF
'to come out'

Latin *insul-arum*

island-GEN.PL
'of the islands'

There are many other rules available at the Leipzig website, as well as lists of suggested abbreviations for various inflectional meanings. The student morphologist should keep in mind, though, that until recently there were no agreed upon conventions for glossing, and one is apt to find all sorts of variations on abbreviations and glossing conventions in articles and grammars about different languages. Luckily, most authors include a list of the abbreviations they use for the reader to consult!

6.2.5 Tense and Aspect

Tense and **aspect** are inflectional categories that usually pertain to verbs. Both have to do with time, but in different ways.

Tense refers to the point of time of an event in relation to another point – generally the point at which the speaker is speaking. In **present** tense the point in time of speaking and of the event spoken about are the same. In **past** tense the time of the event is before the time of speaking. And in **future** tense the event time is after the time of speaking. This can be represented schematically as in (13), where S stands for the time of speaking and E for the time of the event:

- (13) *Present* S = E
 Past E before S
 Future S before E

In English, we mark the past tense using the inflectional suffix *-ed* on verbs (*walked*, *yawned*), but there is no inflectional suffix for future tense. Instead, we use a separate auxiliary verb *will* to form the future tense (*will walk*, *will scream*). The use of a separate word to form a tense is called **periphrastic marking**. Strictly speaking, periphrastic marking is a matter of syntax rather than morphology. Unlike English, Latin marks both past and future inflectionally, that is, by means of morphology on the verb:

- (14) *Present* amō 'I love'
 Past amāvī 'I loved'
 Future amābō 'I will love'

Past, present, and future are not the only possible tenses; some languages distinguish several kinds of past tense and several kinds of future tense, based on how close or distant the event spoken about is from the time of speaking. For example, in Washo (a language isolate), there are four different past tenses and three different future tenses (Mithun 1999: 152–3):

- (15)
- | | |
|-------|---|
| -lul | distant past, before the lifetime of the speaker |
| -gul | distant past, but within the speaker's lifetime |
| -ay? | intermediate past, earlier than the same day, but not extremely distant |
| -leg | recent past, earlier on the same day or during the previous night |
| -áša? | near future, from the point of speaking until about an hour from that point |
| -ti? | intermediate future, after a short lapse of time (usually later in the day) |
| -gab | distant future, following day or later |

Tense on Nouns?

It may seem odd to think of putting tense marking on nouns; we don't do it in English, and it probably isn't done in any of the languages you've studied. But it's not uncommon in native languages of North America. For example, in Central Alaskan Yup'ik (Eskimo-Aleut), both past tense and future tense can be marked on nouns (Mithun 1999: 154):

ikamraqa	'my sled'
ikamralqa	'my former sled'
ikamrarkaqa	'my sled to be'
nuliaqa	'my wife'
nulialqa	'my late/ex-wife'
nuliarkaqa	'my wife to be'

A past tense 'sled' might be a pile of junk in a shed, or something I used to own, now owned by someone else. If your 'wife' is past tense, she might be dead, or you might be divorced. A future tense 'sled' might also be a pile of material yet to be assembled, or something you might buy or get as a present. Your future tense 'wife' is your fiancée.

Aspect is another inflectional category that may be marked on verbs. Rather than showing the time of an event with respect to the point of speaking, aspect conveys information about the internal composition of the event or "the way in which the event occurs in time" (Bhat 1999: 43).

One of the most frequently expressed aspectual distinctions that can be found in the languages of the world is the distinction between perfective and imperfective aspect. With **perfective** aspect, an event is viewed as completed; we look at the event from the outside, and its internal structure is not relevant. With **imperfective** aspect, on the other hand, the event is viewed as ongoing; we look at the event from the inside, as it were. English isn't the best language with which to illustrate this distinction, as tense and aspect are not completely distinct from one another, but I can give you a rough example from English. In English, when we say *I ate the apple*, we not only place the action in the past tense, but also look at it as a completed whole. But if we say *I was eating the apple*, although the action is still in the past, we focus on the event as it is progressing. The Iroquoian language Seneca has a much clearer distinction between perfective and imperfective aspect (Mithun 1999: 165):

- (16) *Perfective* ɔkáhta?t 'I got full'
 Imperfective akáhta?s 'I get full, I'm getting full'

Other forms of aspect focus on particular points in an event. **Inceptive** aspect focuses on the beginning of an event. **Continuative** aspect focuses on the middle of the event as it progresses, and **completive** on the end. We can illustrate these aspects with the following sentences from the Tibeto-Burman language Manipuri (Bhat 1999: 52):

- (17) a. *Inceptive*
 məhak-nə phu-gət-li
 he-NOM beat-start-NON.FUT
 'He began to beat it (and would continue to do so)'
 b. *Continuative*
 tombə layrik pa-rì
 Tomba book read-CONT
 'Tomba is reading the book'
 c. *Completive*
 yumthək ədu yu-rəm-mì
 roof that leak-COMP-CONT
 'That roof had been leaking (but not any more)'

A third category of aspectual distinction can be called **quantificational**. Quantificational aspectual distinctions concern things like the number of times an action is done or an event happens – once or repeatedly – or how frequently an action is done. Among the quantificational aspects are **semelfactive**, **iterative**, and **habitual aspects**. Actions that are done just once are called semelfactive. The Athapaskan language Koyukon has a special verb stem for actions that are done just once (Mithun 1999: 168), and the Eskimo-Aleut language West Greenlandic³

3. Now more commonly known by the autonym Kalallisut.

has suffixes that express iterative aspect for something that is done repeatedly and habitual aspect for something that is usually or characteristically done (Fortescue 1984: 280–2):

- (18) a. *Koyukon semelfactive*
 yeeltleʃ
 ‘she chopped it once, gave it a chop’
- b. *West Greenlandic iterative*
 quirsur-tar-puq
 cough-ITERATIVE-3rd.SG.INDIC
 ‘He coughed repeatedly’
- c. *West Greenlandic habitual*
 qimmi-t qilut-tar-put
 dog.PL bark-HABITUAL-3rd.INDIC
 ‘Dogs bark’

There are other sorts of aspectual distinctions that can be made as well, but these illustrate at least the main types of aspect that can be found in the languages of the world.

English can make some of the aspectual distinctions mentioned above, but it does so periphrastically, using a combination of extra verbs, prepositions, and adverbs to convey such nuances of meaning. In other words, many of these aspectual differences are expressed lexically rather than inflectionally:

- | | | |
|------|---------------------|---------------------------|
| (19) | <i>Inceptive</i> | She began to walk. |
| | <i>Habitual</i> | She always/usually walks. |
| | <i>Continuative</i> | She keeps on walking. |
| | <i>Iterative</i> | She reads over and over. |

Tense and aspect can be and frequently are combined in language, so, for example, it is possible to speak of an event that is ongoing in the past or the future. As mentioned above, tense and aspect are often combined in English. The past tense in English is also typically perfective: when we say *she walked* we generally speak of an event that is conceived of as completed. But when we use the past progressive, as in *she was walking*, we are talking about something that happened in the past, but which we are thinking of as ongoing. The present tense in English can be used to signal the present moment (*At this very moment, a dog barks*), or also to convey habitual aspect (*Dogs bark*).

6.2.6 Voice

Voice is a category of inflection that allows different noun phrases to be focused in sentences. In the **active** voice in a sentence with an agent and a patient,⁴ the agent is focused by virtue of being the subject of the sentence:

- (20) The cat chased the mouse.

4. See Section 8.2, for a discussion of the terms **agent** and **patient**.

But in the **passive** voice the patient is the subject of the sentence, and it gets the focus:

(21) The mouse was chased (by the cat).

In English the passive is expressed periphrastically by a combination of the auxiliary verb *be* plus the past participle, in example (20) *chased*, but in Latin active and passive forms of the verbs are distinguished by inflectional suffixes:

(22)	Active					
	singular	1st	amō	plural	1st	amāmus
		2nd	amās		2nd	amātis
		3rd	amat		3rd	amant
	Passive					
	singular	1st	amor	plural	1st	amāmur
2nd		amāris	2nd		amāminī	
3rd		amātur	3rd		amantur	

There is a great deal more to be said about voice distinctions like active and passive, and we will return to them in some detail in [Chapter 8](#).

6.2.7 Mood and Modality

The inflectional categories of **mood and modality** have to do with a range of distinctions that include signaling the kind of **speech act** in which a verb is deployed. Speech acts are classically defined as things we can do with words: for example, making a statement, asking a question, or giving a command. Languages often have three moods: **declarative** for making ordinary statements, **interrogative** for asking questions, **imperative** for giving commands. But some languages can have other moods as well, for example, expressing a speaker's attitude about a statement, including whether it is necessary, possible, or certain.

Tonkawa (a language isolate) had eight suffixes signaling different moods/modalities ([Mithun 1999: 171](#)). Mood suffixes are shown in bold:

(23) Declarative	naxadjganaw- o	'o.'	'I married'
Assertive	do-nan- a	'a	'He lies!'
Exclamatory	'awac'a-la hedoxa- gwa		'The meat is all gone!'
Interrogative	yaxa-'ga?		'Did you eat?'
Intentive	heul- a -ha'a		'I shall catch him'
Imperative	'andjo- u		'Wake up!'
Potential	ya-dj-'a-n'ec		'I might see him'
Exhortative	hama'amdo-xa-dew- e	l	'Let him be burned up'

Another interesting distinction in mood/modality is the **realis/irrealis** distinction that is marked in some Native American languages. If the **realis** form is used, the speaker means to signal that the event is actual, that it has happened or is happening, or is directly verifiable by perception. The **irrealis** form, in contrast, signals something that can be imagined or thought.

In English, we have no special inflection that signals mood; questions are formed using syntactic means and intonation, imperatives deploy the uninflected verb stem without any special endings. We have a remnant of a **subjunctive** mood, which appears in counterfactual sentences (i.e., sentences expressing something contrary to fact) like *If I were an aardvark, I'd eat insects*; in such sentences the subjunctive verb form is the same as the plural form of the verb (so in the sentence just given, we have *were* rather than *was*).

6.2.8 Evidentiality and Mirativity

In languages that mark the inflectional category of **evidentiality**, speakers must specify in every sentence what the source of information is, whether firsthand knowledge because the speaker has seen or heard it, whether it comes from someone else, whether it is assumed as common knowledge. For example, in Wintu (Wintuan) the suffix *-nt^{he}* signals something that the speaker knows through sensory experience, like seeing, hearing, touching or tasting. The suffix *-ke* indicates that the source of information is hearsay and *-re* indicates something that the speaker has inferred from various clues but does not know by direct personal experience (Mithun 1999: 183–4). Central Pomo (Pomoan) has several clitics that express the source of information; Mithun points out (1999: 181) that theoretically Central Pomo speakers could just say *čhémul* to mean ‘it rained’, but it’s not likely that they would say this. Instead, we might hear one of the five forms in (24), where the circumstances in which any of the forms might be used is given in parentheses:

- (24) *čhémul*=ʔma ‘it rained’ (it’s an established fact)
 čhémul=ya ‘it rained’ (I was there and saw it)
 čhémul=ʔdo· ‘it rained’ (I was told)
 čhémul=nme· ‘it rained’ (I heard the drops on the roof)
 čhémul=ʔka ‘it rained’ (everything is wet)

Of course, in English we could say the same thing by adding an extra clause (like ‘I heard that it rained’) or an adverb (‘supposedly it rained’), but we do not have to express the source of our knowledge as a speaker does in Wintu or Central Pomo. Aikhenvald (2004: 17) estimates that only about 25 percent of the world’s languages express evidentiality as an inflectional category, so it is quite rare. It occurs in a number of Native American languages as well as languages spoken in South America, the Caucasus region, and in certain parts of the Tibeto-Burman language family.

Mirativity is a related category of inflection that marks information as surprising or unexpected. You can see the contrast between a sentence without the mirative marker in (25a) and with it in (25b) from the Nakh-Dagestanian language Chechen (Peterson 2021: 145):

- (25) a. Zaara j-iena
 Zara j-come.PERF
 ‘Zara has come’ [and she is still here; I expected her to come]

- b. Zaara j-iena-q
 Zara J-COME.PERF-MIR
 'Zara has come!' [I didn't expect her to come]

In these sentences, the subject and verb are the same, but the extra marker on the verb expresses the speaker's surprise at the unexpected event.

Challenge

Browse the catalogue or shelves of your university library for grammars of unfamiliar languages (quite a few are available as e-books these days!). Find a grammar of a language that you've never heard of before and see what kinds of inflection it has, if any. Does it inflect nouns? Is there a case system? If so, what kind? What kind of verbal inflections do you find? Do you find any distinctions that we failed to cover in our survey, or other kinds of inflection that strike you as interesting? Share your findings with your classmates.

6.3 Inflection in English

6.3.1 What We Have

As we've seen in passing in the sections above, English is a language that is quite poor in inflection. The distinction between singular and plural is marked on nouns:

- (26) *Singular* cat, mouse, ox, child
 Plural cats, mice, oxen, children

English has only a tiny bit of case marking on nouns: it uses the morpheme *-s* (orthographically *-s* in the singular, *-s'* in the plural) to signal possession, the remnant of the genitive case. Pronouns, however, still exhibit some case distinctions that are no longer marked in nouns:

- (27) *Nouns*
- | | | | |
|--------------------------------|----------|------------|--|
| <i>singular non-possessive</i> | mother | child | |
| <i>singular possessive</i> | mother's | child's | |
| <i>plural non-possessive</i> | mothers | children | |
| <i>plural possessive</i> | mothers' | childrens' | |
- Pronouns*
- | | | | |
|----------------------------|-----|------|-------------|
| <i>singular subject</i> | I | you | he/she/it |
| <i>singular object</i> | me | you | him/her/it |
| <i>singular possessive</i> | my | your | his/her/its |
| <i>plural subject</i> | we | you | they |
| <i>plural object</i> | us | you | them |
| <i>plural possessive</i> | our | your | their |

In verbs, number is only marked in the third person present tense, where *-s* signals a singular subject. As we've seen, English verbs inflect for past tense, but not for future, and there are two participles (present

with *-ing* and past with *-ed*) that together with auxiliary verbs help to signal various aspectual distinctions:

(28)	<i>Verbs</i>	
	<i>3rd person sg. present</i>	walks, runs
	<i>all other present tense forms</i>	walk, run
	<i>past tense</i>	walked, ran
	<i>progressive</i>	(be) walking, running
	<i>past participle</i>	(have) walked, run

Distinctions in aspect and voice are expressed in English through a combination of auxiliary choice and choice of participle. The **progressive**, which expresses, among other things, ongoing actions, is formed with the auxiliary *be* plus the present participle:

(29)	<i>Present progressive</i>	I am mowing the lawn.
	<i>Past progressive</i>	I was mowing the lawn.
	<i>Future progressive</i>	I will be mowing the lawn.

The **perfect** (note that the perfect is not the same as the perfective, which we discussed above) expresses something that happened in the past but still has relevance to the present. This is signaled in English with the past participle and a form of the auxiliary *have*:

(30)	<i>Perfect</i>	I have eaten the last piece of blueberry pie.
------	----------------	---

The passive voice in English is formed with the past participle as well, but the auxiliary *be* is used instead of *have*:

(31)	<i>Passive</i>	I was followed by a voracious weasel.
------	----------------	---------------------------------------

It is, of course, possible to combine various auxiliaries and participial forms to express tense/aspect distinctions that are quite complex, as in, for example, the past perfect progressive passive sentence *I had been being followed by a voracious weasel*.

As you can see in (26)–(31), English has both regular and irregular inflections. All of our regular inflections are suffixal, but irregular forms are often formed by internal stem change (ablaut and umlaut) or by a combination of internal stem change and suffixation. Examples of irregular forms are given in (32):

(32)	<i>a. Irregular noun plurals</i>		
	foot	feet	
	mouse	mice	
	ox	oxen	
	child	children	
	alumnus	alumni	
	datum	data	
	<i>b. Irregular verb forms</i>		
	sing	sang	sung
	sit	sat	sat
	swing	swung	swung

write	wrote	written
hold	held	held
tell	told	told
bring	brought	brought

We could no doubt think of more examples as well. It is often said that the irregular plurals and past tenses in English form **closed classes**; that is, they constitute a fixed list from which particular forms can be lost, but to which no new forms can be added. The regular plural and past tense endings are considered **default endings**. In other words, when a new noun is added to English, its plural is formed with -s and when a new verb is added, its past tense is formed with -ed. So if we borrow a noun from another language or coin a completely new noun, their plurals will be formed with the regular suffix (*fajitas*, *wugs*). Similarly for past tenses of new verbs (*googled*).

Why do we have irregular forms? In some cases, they are the remnants of ways of forming the plural or past tense that we no longer have today. We saw in [Chapter 5](#) that plurals like *foot* ~ *feet* and *mouse* ~ *mice* are the remnants of a rule of umlaut that was lost at the earliest stages of English. The irregular verbs are also a remnant of a way of inflecting verbs that goes all the way back to the Germanic ancestor of English and even beyond that to the way of inflecting verbs in proto-Indo-European. The details of how that kind of inflection worked need not concern us here, except to say that it involved internal vowel changes (ablaut). Suffice it to say that that form of inflection is no longer used to inflect new verbs in English.

Challenge

Is it really true that we can never form new irregular verbs? Here's an experiment that you and your classmates can do. Ask five friends to fill in the past tense of each nonsense verb in the following sentences:

Max likes to *plite*. Yesterday he _____ for two full hours.
 Zelda *glings* very well. Years ago, her mother _____ professionally.
 Sometimes we *trell* all night. Last weekend we _____ for two whole days.

Compile your data with that of your classmates and try to see if there are any patterns you can explain.

6.3.2 Why English Has So Little Inflection

You might wonder why English has so little inflection. In fact, if you study the history of English, you'll find that at one time English had quite a bit more inflection than it now has. The earliest form of English, Old English, was spoken from about 450 to 1100 CE. Old English had three genders – masculine, feminine, and neuter – and a case system with four cases – nominative, accusative, dative, and

genitive; see (33). Articles and adjectives agreed with nouns in case and number, as (34) shows:

(33) *Old English nouns*

	<i>Masculine</i>	<i>Feminine</i>	<i>Neuter</i>
<i>Singular</i>			
Nom.	stān ‘stone’	giefu ‘gift’	scip ‘ship’
Acc.	stān	giefe	scip
Gen.	stānes	giefe	scipes
Dat.	stāne	giefe	scipum
<i>Plural</i>			
Nom.	stānas	giefa	scipu
Acc.	stānas	giefa	scipu
Gen.	stāna	giefa	scipe
Dat.	stānum	giefum	scipum

(34)	sē	gōdan	stān
	the.MASC.NOM	good.MASC.NOM	stone.NOM
	sēo	gōde	giefu
	the.FEM.NOM	good.FEM.NOM	gift.NOM
	ðæt	gōde	scip
	the.NEUT.NOM	good.NEUT.NOM	ship.NOM

Verb inflection was also more complex in Old English than in Modern English. In Old English, verbs were inflected for person and number, and were different for present tense and past tense in both the indicative and the subjunctive. In addition, some verbs were **strong verbs**, which means that they show internal stem change – specifically ablaut – in some forms. For example, in the strong verb *drīfan* ‘to drive’ the present tense always has a long [i] vowel, but the past has the vowel [a] in the first and third person singular and a short [i] in the plural form and the second person singular. Other verbs were called **weak verbs**; these inflected using suffixes rather than ablaut. As you can see with the weak verb *dēman* ‘to judge’ in (35), the vowel of the stem never changes, but in the past tense there is always a suffix *-d(e)*:

(35)		<i>Strong</i>	<i>Weak</i>
		drīfan ‘to drive’	dēman ‘to judge’
	<i>Present</i>		
	<i>singular</i>		
	1st	drīfe	dēme
	2nd	drīf(e)st	dēm(e)st
	3rd	drīf(e)ð	dēm(e)ð
	<i>plural</i>		
	1st	drīfað	dēmað
	2nd	drīfað	dēmað
	3rd	drīfað	dēmað
	<i>Past</i>		
	<i>singular</i>		
	1st	drāf	dēmde
	2nd	drife	dēmdes(t)
	3rd	drāf	dēmde

<i>plural</i>	1st	drifon	dēmdon
	2nd	drifon	dēmdon
	3rd	drifon	dēmdon

Why did English lose all this inflection? There are probably two reasons. The first one has to do with the stress system of English: in Old English, unlike modern English, stress was typically on the first syllable of the word. Ends of words were less prominent, and therefore tended to be pronounced less distinctly than beginnings of words, so inflectional suffixes tended not to be emphasized. Over time this led to a weakening of the inflectional system. But this alone probably wouldn't have resulted in the nearly complete loss of inflectional marking that is the situation in present day English; after all, German – a language closely related to English – also shows stress on the initial syllables of words, and nevertheless has not lost most of its inflection over the centuries.

Some scholars attribute the loss of inflection to language contact in the northern parts of Britain. For some centuries during the Old English period, northern parts of Britain were occupied by the Danes, who were speakers of Old Norse. Old Norse is closely related to Old English, with a similar system of four cases, masculine, feminine, and neuter genders, and so on. The actual inflectional endings, however, were different, although the two languages shared a fair number of lexical stems. For example, the stem *bōt* meant 'remedy' in both languages, and the nominative singular in both languages was the same. But the nominative plural in Old English was *bōta* and in Old Norse *bótaR*.⁵ The form *bóta* happened to be the genitive plural in Old Norse. Some scholars hypothesize that speakers of Old English and Old Norse could communicate with each other to some extent, but the inflectional endings caused confusion, and therefore came to be de-emphasized or dropped. One piece of evidence for this hypothesis is that inflection appears to have been lost much earlier in the northern parts of Britain where Old Norse speakers cohabited with Old English speakers, than in the southern parts of Britain, which were not exposed to Old Norse. Inflectional loss spread from north to south, until all parts of Britain were eventually equally poor in inflection (O'Neil 1980; Fennell 2001: 128–9).

6.4 Paradigms

If you've ever studied a foreign language – French, Latin, German, Russian – you probably know at least intuitively what a **paradigm** is. A paradigm consists of all of the different inflectional forms of a particular lexeme or class of lexemes. Each distinct form of a lexeme exhibits a specific combination of the inflectional properties that are expressed in that language. For convenience, you can think of a paradigm as a kind of table or grid with cells, one for each inflected form for a given lexeme. For example, in (33) and (35) above, I've shown you paradigms for the

5. The *R* here is a runic character with a phonetic value close to [z].

Old English nouns ‘stone’, ‘gift’, and ‘ship’, and for the verbs ‘to drive’ and ‘to judge’; these paradigms show the various endings and stem changes that are exhibited by nouns of different genders in the singular and plural in different cases, and in the present and past of strong and weak verbs in different persons. Traditionally the paradigm of a noun or adjective was called its **declension** and that of a verb its **conjugation**.

6.4.1 Inflectional Classes

Within a language, not all nouns or verbs may inflect in exactly the same way. All members of a particular category will typically make the same inflectional distinctions, for example, exhibiting case, number, or tense; but the actual forms for particular cases, numbers, or tenses might differ from one group of nouns or verbs to another. These different inflectional subpatterns are called **inflectional classes**. In Latin, for example, nouns generally belong to one of five inflectional classes that differ to some extent in their inflectional suffixes:⁶

(36)	1	2	3	4
	‘star’	‘servant’	‘father’	‘hand’
<i>Stem form</i>	stellā (F.)	servo (M.)	patr (M.)	manu (F.)
<i>Singular</i>				
Nom.	stella	servus	pater	manus
Gen.	stellae	servī	patris	manūs
Dat.	stellae	servō	patrī	manuī
Acc.	stellam	servum	patrem	manum
Abl.	stellā	servō	patre	manū
<i>Plural</i>				
Nom.	stellae	servī	patrēs	manūs
Gen.	stellārum	servōrum	patrum	manuum
Dat.	stellīs	servīs	patribus	manibus
Acc.	stellās	servōs	patrēs	manūs
Abl.	stellīs	servīs	patribus	manibus

Nouns that belong to the first inflectional class, traditionally called first declension nouns, are usually feminine, and their stems always end in a long *-ā*. Second declension nouns are typically masculine or neuter and have stems ending in *-o*. In the third declension, nouns may be of any gender and stems typically end in a consonant. And so on.

Latin verbs fall into four inflectional classes or conjugations. Each conjugation is characterized by a particular vowel, called the **theme vowel**, which has no meaning, but is suffixed to the verb root to form a stem.

(37)	1	2	3	4
	‘love’	‘warn’	‘say’	‘hear’
<i>Root</i>	am	mon	dic	aud
<i>Stem (root + theme vowel)</i>	am + ā	mon + ē	dic + i	aud + ī

6. This is a slight simplification, because some of the Latin inflectional classes also have subclasses.

<i>Present</i>					
singular	1st	am+ō	mon+e+ō	dic+ō	aud+i+ō
	2nd	am+ā+s	mon+ē+s	dic+i+s	aud+ī+s
	3rd	am+a+t	mon+e+t	dic+i+t	aud+i+t
plural	1st	am+ā+mus	mon+ē+mus	dic+i+mus	aud+ī+mus
	2nd	am+ā+tis	mon+ē+tis	dic+i+tis	aud+ī+tis
	3rd	am+a+nt	mon+e+nt	dic+unt	aud+i+unt

The person and number endings are attached directly to the root in the first person singular of the first and third conjugations, and otherwise to the stem, which consists of the root plus the theme vowel.

6.4.2 Suppletion, Syncretism, Overabundance, and Defectiveness

Suppletion, **syncretism**, **overabundance**, and **defectiveness** are terms that refer to relationships between inflected forms in a paradigm.

Suppletion occurs when one or more of the inflected forms of a lexeme is built on a base that bears no relationship to the base of other members of the paradigm. Consider, for example, the verb *go* in English. In the present tense the base is *go*, of course: *I, you, we, they go; he/she/it goes*. The progressive participle is *going*, and the past participle *gone*, both built on the base *go* as well. The past tense of *go*, however, is a suppletive form *went* – that is, a base that is completely different from that of all the other forms. The Latin verb *ferō* is notorious for its suppletive forms. Its present stem is *fer* (so, for example, the first person plural form is *ferimus*). However, its past tense forms are built on the stem *tul*, and some of its participles are built on yet a third stem *lāt*.

Syncretism is another relationship we can find between the members of a paradigm, specifically one in which the same form fills two or more ‘cells’ in our inflectional grid or table. Consider, for example, the Old English verb paradigm we looked at earlier (repeated here for convenience):

(38)	<i>Present</i>		drīfan ‘to drive’	dēman ‘to judge’
	singular	1st	drīfe	dēme
		2nd	drīf(e)st	dēm(e)st
		3rd	drīf(e)ð	dēm(e)ð
	plural	1st	drīfað	dēmað
		2nd	drīfað	dēmað
		3rd	drīfað	dēmað

Although there are distinct forms for first, second, and third person in the singular, the same form is used for all three persons in the plural. These forms display syncretism. Another, more complex example of syncretism comes from the noun paradigms of a West Slavic language, Sorbian, spoken in parts of Germany (Baerman 2007):

(39)		<i>Singular</i>	<i>Dual</i>	<i>Plural</i>
	<i>Nominative</i>	žona	žone	žony
	<i>Accusative</i>	žonu	žone	žony
	<i>Genitive</i>	žony	žonow	žonow

<i>Locative</i>	žonje	žonomaj	žonach
<i>Dative</i>	žonje	žonomaj	žonam
<i>Instrumental</i>	žonu	žonomaj	žonami

The chart (39) illustrates the paradigm for the noun ‘woman’. You can see that in the singular the accusative and instrumental cases are syncretic – they have exactly the same inflectional form – as are the dative and locative cases.⁷ In the dual, nominative and accusative are syncretic, as are locative, dative, and instrumental. And in the plural, nominative and accusative are syncretic as well. Morphologists are interested in seeing if there are any patterns to syncretism across languages – for example, is it more typical in languages for plural forms of verbs to be syncretic than singular forms? Or is it more common for nominative and accusative forms of nouns to be syncretic than, say, nominative and dative? As yet there is no definitive answer to these questions.

Normally each cell in a paradigm’s table contains one and only one form. But occasionally in languages there is more than one form that can fill a given cell. This situation is referred to as **overabundance** (Thornton 2011, Stump 2016). For example, the normal situation for English is for each verb to have a single past tense form. The past tense of *seem* is *seemed*. The past tense of *weep* is *wept*. But for many speakers of English the past tense of the verb *dream* can be either *dreamed* or *dreamt*. This is what’s known as overabundance.⁸

Defectiveness occurs when there are one or more cells in the paradigm that unaccountably are filled by no forms at all. For example, apparently in French the verb *traire* ‘to milk’ has present and future forms, but does not have any preterite forms (Stump 2016: 160). Another example is the verb *beware* in English, which we can use in imperative sentences (for example, *Beware of vampires!*), but which is almost never attested in any other inflectional form. The *OED* shows a couple of examples, the most recent being in the nineteenth century of the forms *bewared* or *bewaring*, but I doubt we’d see these today.

6.5 Inflection and Productivity

It is often said that inflection differs from derivation in terms of productivity. We saw in Chapter 4 that some rules of word formation are more productive than others. There are derivational affixes in English, as we saw, that are quite dead or nearly so, others that are relatively productive, and some that are fully productive. In contrast, rules of inflection are almost always fully productive: every verb in English, for example, has a progressive form with the suffix *-ing*, and just about every verb can

7. The instrumental case is used to mark ‘instruments’; for example, in a sentence like *I cut the bread with a knife*, the noun *knife* would be in the instrumental case. The locative case is used to mark locations; for example, the phrase *in the trees* would be locative in *The birds were singing in the trees*.

8. Note that different dialects have different past tense forms for a verb, say *dived* in one dialect and *dove* in another. Here we mean that an individual speaker uses both forms more or less randomly.

form a past tense. I say “just about every verb” because there are occasional verbs that native speakers of English (at least of American English) are highly reluctant to use in the past tense: for example, I can use the verb *forgo/forego* in the present tense (*I forgo dessert most nights*), but for the past tense *forwent* sounds too odd for most people to use either in spoken or written form (*??I forwent dessert last night*), and there’s no alternative. Certainly not *forgoed*!

Challenge

Are there any other verbs in English that you can’t use in the past tense? Note that what I mean here are verbs for which you might find past tense forms in the dictionary, but would never actually use yourself. Find a list of irregular verbs in English, either in a dictionary, a grammar book, or online, and see if you can find some. Then compare your list of “no past tense” verbs with your classmates’. Is there any generalization you can make about the verbs for which we have no past tenses?

6.6 Inherent versus Contextual Inflection

Consider the Latin phrases in (40):

(40)	bonus	puer
	good-MASC.SG.NOM	boy-MASC.SG.NOM
	‘good boy’	
	bona	puella
	good-FEM.SG.NOM	girl-FEM.SG.NOM
	‘good girl’	
	bonum	dōnum
	good-NEUT.SG.NOM	gift-NEUT.SG.NOM
	‘good gift’	

In each of these phrases the adjective and noun agree in number, gender, and case. As you can see, the adjective ‘good’ can occur in any of the three genders, depending on the context in which it occurs. The nouns ‘boy’, ‘girl’, and ‘gift’ are always, however, respectively masculine, feminine, and neuter nouns. In other words, the inflectional category of gender is **inherent** in nouns, whereas it is **contextual** in adjectives.

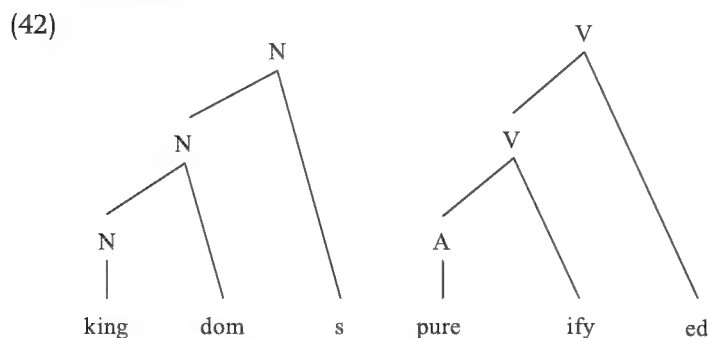
Put more generally, **contextual inflection** is inflection that is determined by the syntactic construction in which a word finds itself, whereas **inherent inflection** is inflection that does not depend on the syntactic context in which a word finds itself. Number is inherent in nouns and pronouns, as is gender. But the case of a noun always depends on its syntactic context. On the other hand, tense and aspect are inherent in verbs, but person and number depend on the nouns or pronouns with which a verb occurs in a sentence. So number can be **contextual** for one category (verbs), but **inherent** for another (nouns).

6.7 Inflection *versus* Derivation Revisited

We have now looked in some detail both at inflection and at derivation (or lexeme formation, more generally). Remember that we distinguished the two sorts of morphology in the following ways:

(41)	<i>Inflection</i>	<i>Derivation</i>
	never changes category	sometimes changes category
	adds grammatical meaning	often adds lexical meaning
	is important to syntax	produces new lexemes
	is usually fully productive	can range from unproductive to fully productive

One more thing we can add to these differences is that in words that have both inflectional and derivational affixes, the derivational affixes almost always occur inside the inflectional ones. Remember from [Chapter 3](#) that words are formed like onions, with successive layers added one by one. In these word structures, derivational affixes go closer to the base (or root or stem) than inflectional affixes do. For example, in English we can have words like *kingdoms* where the noun suffix *-dom* attaches to its base before the plural suffix *-s*, or *purified* where the verb-forming suffix *-ify* attaches to the adjective *pure* before the past tense suffix *-ed* is added:



It's not possible in English to attach a plural or past tense suffix and then a derivational suffix; we never form words like **kingdom* or **walkeder*.

Challenge

We have seen that in words that have both derivational affixation and inflection derivation occurs 'inside' inflection. Now consider compounding. Is it possible for the first element in a compound in English to bear an inflection, like a plural or a possessive suffix? You'll have to think carefully here. Collect relevant examples and see if you can make up some yourselves. What do you think: can inflection occur 'inside' compounds?

We've seen that the differences between inflection and derivation are usually quite clear, in terms of function, meaning, and position in word structures. In most cases, when we look at the morphology of languages it's not too hard to distinguish inflection from derivation. There are cases, however, where the distinction doesn't seem so clear-cut. One puzzling case is that of nouns in Fula (Lieber 1987: 74; Arnott 1970: 75). Each noun in Fula belongs to up to seven different noun classes. One of those classes is a singular class, another a plural class, and the remainder are classes for singular and plural diminutives, pejorative diminutives (meaning something like 'nasty little X'), augmentatives, and pejorative augmentatives. The various noun class forms for 'monkey' are shown in (43):

(43) *Fula (Niger-Congo)*

	<i>waa</i>	'monkey'
11	<i>waa-ndu</i>	singular
25	<i>baa-dfi</i>	plural
3	<i>baa-ngel</i>	diminutive singular
5	<i>baa-ngum</i>	pejorative diminutive singular
6	<i>mbaa-kon</i>	diminutive plural
7	<i>mbaa-nga</i>	augmentative singular
8	<i>mbaa-ko</i>	augmentative plural

As (43) illustrates, noun class is marked in Fula not only by different suffixes, but also by mutation of the initial consonant of the noun stem. So in class 11, for example, the initial consonant of the stem is a continuant [w], in classes 25, 3, and 5, the initial consonant is a stop [b] that corresponds in point of articulation to the continuant, and in classes 6, 7, and 8 it is a prenasalized stop [mb], again corresponding in point of articulation. Singular and plural, of course, are number distinctions, and thus belong to the realm of inflection. Augmentatives and diminutives, however, are expressive morphology (look back to Chapter 3, Section 3), and are usually considered derivational. The whole list in (43) has the look of a paradigm, though, and even more perplexing, adjectives and articles agree with Fula nouns, even in the noun classes that form augmentatives and diminutives. Agreement, of course, is a hallmark of inflection. The point here is that it's not at all clear in the case of Fula whether to count augmentatives and diminutives as inflectional or derivational, or to concede that in this particular case the distinction just doesn't make sense.

An equally perplexing case can be found in Sesotho. Like Fula, Sesotho has noun classes. Classes 1 and 2 are classes that contain human nouns. Interestingly, agent nouns can be derived from verbs by putting them into classes 1/2:

(44) *Sesotho (Niger-Congo)* (Demuth 2000: 278)

<i>Infinitive of verb</i>		<i>Corresponding agent noun (class 1)</i>	
<i>ho-pheha</i>	'to cook'	<i>mo-phehi</i>	'cook'
<i>ho-ruta</i>	'to teach'	<i>mo-ruti</i>	'teacher'

Formation of agent nouns from verbs certainly looks like derivation. Yet in Sesotho, as in Fula, articles and adjectives must agree with the nouns they modify, in that they must bear the class prefixes corresponding to those nouns. And agreement, of course, is part of inflection.

We leave this issue unresolved here – although it may not seem particularly satisfying to a beginning student of morphology to know that the distinction between inflection and derivation is not always crystal clear, it is precisely cases like these that most intrigue morphologists!

6.8 How To: More Morphological Analysis

In this chapter, I have tried to expose you to some of the sorts of inflectional distinctions that you might expect to find outside of English. Reading about inflectional morphology is not the same as analyzing it though. As a student morphologist you should be ready to figure out what sorts of inflectional distinctions are expressed in an unfamiliar language and how they are expressed. In this section, I will simulate such an experience for you, and take you through the process of trying to analyze the inflectional morphology of a language you otherwise know little or nothing about. Our example comes from Yimas, (Foley 1991: 217):

(45) *Yimas verb forms (Ramu-Lower Sepik)*

- | | | |
|----|--------------|---|
| a. | nakatay | 'I see him' |
| b. | nantay | 'you see him' |
| c. | nantay | 'he sees him' |
| d. | impakatay | 'I see those two' |
| e. | impantay | 'you see those two' |
| f. | impantay | 'he sees those two' |
| g. | pukatay | 'I see them (more than two)' |
| h. | puntay | 'you see them' |
| i. | puntay | 'he sees them' |
| j. | naŋkratay | 'we two see him' |
| k. | naŋkrantay | 'you two see him' |
| l. | nampitay | 'those two see him' |
| m. | impaŋkratay | 'we two see those two' |
| n. | impaŋkrantay | 'you two see those two' |
| o. | impampitay | 'those two see those (other) two' |
| p. | puŋkratay | 'we two see them (more than two)' |
| q. | puŋkrantay | 'you two see them' |
| r. | pumpitay | 'those two see them' |
| s. | nakaycay | 'we (more than two) see him' |
| t. | nanantay | 'you (more than two) see him' |
| u. | namputay | 'they (more than two) see him' |
| v. | impakaycay | 'we (more than two) see those two' |
| w. | impanantay | 'you (more than two) see those two' |
| x. | impamputay | 'they (more than two) see those two' |
| y. | pukaycay | 'we (more than two) see them (more than two)' |

z.	punantay	'you (more than two) see them (more than two)'
aa.	pumputay	'they (more than two) see them (more than two)'

Analyzing the inflectional system of an unfamiliar language is not very different from analyzing the sort of derivational data we looked at in [Chapter 3](#). What we do to start off is to look at the translations of the forms, and compare forms for areas of overlap.

If you glance even casually at the data in (45), you'll see that I've given you part of a verb paradigm, namely the verb 'to see' in Yimas. You might therefore expect to find something that occurs in every form that would correspond to the root meaning 'see'. Scanning the data, you will see that almost every form except (45s, v, y) ends in the sequence *tay*. The three forms that don't end in *tay* end in *cay*. It would be a good initial guess, then, that the morpheme for 'see' is *tay*. Since *cay* looks a lot like *tay*, we might hypothesize that it also means 'see', although we don't yet know why those three forms are different. At this point, let's set aside the three forms with *cay* to think more about later.

The next thing we notice about the forms in (45) is that they vary in the person and number of the subject: there are examples with first, second, and third person subjects, not only in singular and plural, but also apparently in the dual. So Yimas makes a three-way number distinction in its verb forms. We notice as well that the examples differ from one another in the number of their object (again singular, dual, and plural), although all of the object forms in these examples happen to be in the third person. Given that we already suspect that the verb root comes at the end, we would guess that subject and object are marked on the verb with prefixes. So our next task is to start comparing the various forms that have the same number object or the same person and number subject to see if we can find separate prefixes for subject and object.

Let's start with the object forms. If you look at (45a–c), you'll notice that all three forms begin with *na-*, but that (45a) has *ka-* after the *na-*, whereas in (45b, c), there's an *n-* after *na-*. We can hypothesize, then, that *na-* must mean 'he-object', but we need to check ourselves to make sure that other forms in (45) with third person singular objects begin with *na-*. If we look at (45j, k, l), you'll see that our hypothesis is confirmed, and if you look further down the data, you'll also see *na-* at the beginning of (45s, t, u). Now, looking at (45d, e, f), you'll notice that all three forms begin with *impa-*; in (45d) this is followed by *ka* and in (45e–f) by *n-*. We can guess then that *impa-* must mean third person dual object. Finally (45g, h, i) all begin with *pu-*; again in (45g) *pu-* is followed by *ka-* and in (45h–i) by *n-*. It looks like *pu-* is used for a third person plural object.

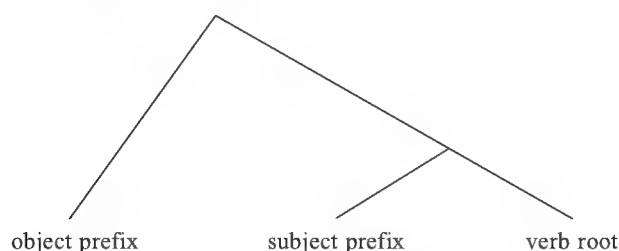
If *na-* is the third person object marker, then what's left over between it and the verb stem must be the subject marker: in (45a, b, c), this would leave *ka-* as the first person singular subject marker, and *n-* for both the second and third person singular subject markers (syncretism!). Now, if you peel off the other object markers *impa-* and *pu-* from the beginnings of the words, and *tay/cay* from the end of the words, what you have left should be all the subject markers. This is what we get:

(46) *Yimas subject markers*

1	sg.	ka-
2	sg.	n-
3	sg.	n-
1	dual	ŋkra-
2	dual	ŋkran-
3	dual	mpi-
1	pl.	kay-
2	pl.	nan-
3	pl.	mpu-

Beginning students might have the impulse to call the subject markers ‘infixes’, because they occur between the object markers and the verb root, but they are not infixes. Remember that we only call an affix an infix if it occurs within another morpheme. What we have here instead is a sequence of two prefixes before the verb root. So the structure of the Yimas verb is:

(47)



A note about the Yimas data: we still haven’t explained why the verb root appears to be *tay* in most verb forms but *cay* in (45s, v, y). With the data I’ve given you, it’s actually impossible to say anything more about these forms. This is a point at which you, as the morphologist studying an unfamiliar language, must go looking for more data in the hope that they will explain what’s going on. So here’s a cautionary note: linguistic data can sometimes be a bit messy. Introductory text books have a tendency to sanitize data so that the student linguist will not be confronted with bits that can’t be explained. But keep in mind that in the real world data are rarely perfectly sanitary. You often see a pattern, but it’s not always perfect. This may be frustrating, but it isn’t a bad thing – every bit of mess – or *apparent* mess – forces us, as linguists, to keep looking for more data and to look more carefully at the data we have.

A final note: if you look carefully in [Foley’s \(1991\)](#) grammar of Yimas, it appears that there is a phonological rule that turns [t] to [c] if it is preceded by [y].⁹ So we were right to guess that *tay* and *cay* are the same morpheme. In [Chapter 9](#), we’ll see more data like this, and learn how to handle them.

Summary

In this chapter we’ve first surveyed quite a few sorts of inflection that can be found in the languages of the world, looking at person, number, gender and noun class, tense and aspect, voice, mood,

9. This is a rough statement of the rule.

modality, and evidentiality. We've looked in some detail at the sorts of inflections that are found in English, and considered the historical reasons why English has relatively little in the way of inflection. We've then looked at paradigms and important relationships between forms in paradigms, such as suppletion and syncretism, and at the distinction between inherent and contextual inflection. We have revisited the distinction between inflection and derivation to see that the line between them can be blurred. And finally, we've looked at a set of data to see how to figure out how the inflection in an unfamiliar language works.

Exercises

- Look at the following data. In each case, identify the form of the morphological rule (i.e., prefixation, suffixation, infixation, reduplication, internal stem change, templatic) and its function (inflection or derivation):

Turkish (Turkic) (Kornfilt 1997: 446)

silâh	'weapon'	silâhlı	'armed person'
at	'horse'	atlı	'horseman'
yaş	'age'	yaşlı	'aged person'
Londra	'London'	Londralı	'person living in London'

Musqueam (Salish) (Suttles 2004: 139)

p'é ^t	'sew'	p'ép'é ^t	'be sewing'
k'wéc	'look'	k'wék'wéc	'be looking'
łás	'fish with a net'	łáłəs	'be fishing with a net'

Hausa (Afro-Asiatic) (Newman 2000: 454)

tsàkō	'chick'	tsàkī	'chicks'
kwādō	'frog'	kwādī	'frogs'
zābō	'guinea-fowl'	zābī	'guinea-fowls'

- In the two columns below, you find verb bases and imperfective forms for a number of verbs in Tagalog (Austronesian) (Schachter and Otnes 1972: 365):

Verb base		Imperfective	
lagyan	'put in/on'	nilalagyan	'is putting in/on'
regaluhan	'give a present to'	nireregaluhan	'is giving a present to'
walisan	'sweep'	niwawalisan	'is sweeping'

Write a rule that shows how the imperfective is formed in Tagalog.

Now consider these data:

Verb base	Imperfective
lagyan	linalagyan
regaluhan	rineregaluhan
walisan	winawalisan

Apparently the imperfective forms in (b) are less preferred, but possible, imperfective forms for the same verbs. Write an alternative rule that accounts for these forms.

Finally, what are the two imperfective forms that you might expect from the verb stem *yapakan* 'step on'?

3. Consider the following data from Swahili (Niger-Congo) (Corbett 1991: 43–4):

- | | | | | |
|----|-----------|---------|---------|------------|
| a. | kikapu | kikubwa | kimoja | kilianguka |
| | 'basket' | 'large' | 'one' | 'fell' |
| b. | vikapu | vikubwa | vitatu | vilianguka |
| | 'baskets' | 'large' | 'three' | 'fell' |

Describe how number marking works in Swahili. On which categories is number marking inherent and on which is it contextual?

4. In Russian, both the noun *student* 'student' and the noun *dub* 'oak' are masculine, but there are slightly different declensions for animate and inanimate nouns. Discuss the paradigms below in terms of the patterns of syncretism they display (data from Corbett 1991: 166):

<i>Singular</i>	<i>student</i>	'student'	<i>dub</i>	'oak'
Nominative	student		dub	
Accusative	studenta		dub	
Genitive	studenta		duba	
Dative	studentu		dubu	
Instrumental	studentom		dubom	
Locative	studente		dube	
<i>Plural</i>				
Nominative	studenty		duby	
Accusative	studentov		duby	
Genitive	studentov		dubov	
Dative	studentam		dubam	
Instrumental	studentami		dubami	
Locative	studentax		dubax	

5. Consider the data below from Syrian Arabic (Afro-Asiatic) (Cowell 1964: 173–4). Segment the words into morphemes, and identify the meanings/functions of the morphemes. What word formation processes are represented in these data? (Hint: Assume that the form for 'he ate' is *ákal*, that this form has neither prefixes nor suffixes, and that a glottal stop always occurs before a vowel-initial stem.)

ʔákal	'he ate'	byaakol	'he eats'
ʔáklet	'she ate'	btaakol	'she eats'
ʔákalu	'they ate'	byaaklu	'they eat'
ʔakált	'you (masc.) ate'	btaakol	'you (masc.) eat'
ʔakálti	'you (fem.) ate'	btaakli	'you (fem.) eat'
ʔakáltu	'you (pl.) ate'	btaaklu	'you (pl.) eat'
kol	'eat! (masc.)'		

6. Below are the Latin verb paradigms for the verbs 'love' and 'warn' in the future and perfect tenses. Identify the morphemes in each form (root, stem, suffixes) and discuss how the future and perfect tenses differ from one another, and how the first and second conjugation verbs differ from one another.

amābō	'I will love'	monēbō	'I will warn'
amābis	'you will love'	monēbis	'you will warn'
amābit	'he/she will love'	monēbit	'he/she will warn'
amābimus	'we will love'	monēbimus	'we will warn'
amābitis	'you (pl.) will love'	monēbitis	'you (pl.) will warn'
amābunt	'they will love'	monēbunt	'they will warn'
amāvī	'I loved'	monuī	'I warned'
amāvistī	'you loved'	monuistī	'you warned'
amāvit	'he/she loved'	monuit	'he/she warned'
amāvimus	'we loved'	monuimus	'we warned'
amāvistis	'you (pl.) loved'	monuistis	'you (pl.) warned'
amāvērunt	'they loved'	monuērunt	'they warned'

7. Dutch makes a distinction between weak and strong verbs. Below are the paradigms for the past tense of a weak verb (*werken* 'to work') and a strong verb (*binden* 'to tie'). Discuss differences in the ways that weak and strong past tenses are formed in Dutch. Are the patterns of syncretism the same or different in the weak and strong forms?

ik werkte	'I worked'	ik bond	'I tied'
jij werkte	'you worked'	jij bond	'you tied'
hij/zij werkte	'he/she worked'	hij/zij bond	'he/she tied'
wij werkten	'we worked'	wij bonden	'we tied'
jullie werkten	'you (pl.) worked'	jullie bonden	'you (pl.) tied'
zij werkten	'they worked'	zij bonden	'they tied'

8. Consider the sets of verbs below from Diegueño (Yuman) (Langdon 1970: 80–7).

- a. a'ap 'to lay down a long object'
- a'kat. 'to cut with a knife'
- a'maʔ 'to sweep'
- a'naɾ 'to lower a long object, to drown'
- a'maɾ 'to cover over a long object, to bury someone'
- a'uɫ 'to lay a long object on top of'
- b. cu'kat. 'to bite off'
- cu'paɾ 'to emit a victory yell'
- cu'kuw 'to bite'
- cu'ya'y 'to hum'
- cu'sip 'to smoke' (e.g., a pipe)
- cu'kʷis 'to chatter (like squirrel)'
- c. tu'kat. 'to cut with scissors or ax, to cut in chunks'
- tu'miɫ 'to hang (small round object)'
- tu'paɾ 'to crack acorns'
- tu'uɫ 'to put on (e.g., a hat)'
- tu'maɾ 'to cover over a small object'
- tu'yum 'to put a round small object in sun'

- (i) Divide the words above into prefixes and roots and try to assign meanings to each morpheme.
- (ii) Is the process you see illustrated here one of inflection or derivation? Give evidence to support your answer.

9. The following data are from Na'vi, a constructed language (conlang) created by the linguist Paul Frommer for the movie *Avatar*. Divide the words below into morphemes and discuss the position of the morphemes with respect to the various verb roots (data from <http://files.learnnavi.org/docs/horenlenavi.pdf>). Note that the verb 'feed' is a compound word.

	eyk	fpak	taron	yomting
	'lead'	'stop'	'hunt'	'feed'
near past	imeyk	fpimak	timaron	yomtining
reflexive	äpeyk	fpäpak	täparon	yomtäping
refl. near past	äpimeyk	fpäpimak	täpimaron	yomtäpiming
ceremonial	uyeyk	fpuyak	taruyon	yomtuying
perf. cerem.	oluyeyk	fpoluyak	tolaruyon	yomtoluying
refl. perf. cerem.	äpoluyeyk	fpäpoluyak	täpolaruyon	yomtäpoluying

10. Consider the data below from Central Guerrero Nahuatl, a Uto-Aztecan language spoken in Mexico (data from Amith and Smith-Stark 1994: 349–51). Assume that the verb base 'love' is *lasola*. Divide the words into morphemes and identify what each morpheme means. Discuss any difficulties you encounter.

a. tine:člasola	'you (sg.) love me'
b. ne:člasola	'he/she/it loves me'
c. nane:člasolan	'you (pl.) love me'
d. ne:člasolan	'they love me'
e. timiçlasola	'I love you (sg.)'
f. miçlasola	'he/she/it loves you (sg.)'
g. timiçlasolan	'we love you (sg.)'
h. miçlasolan	'they love you (sg.)'
i. niklasola	'I love him/her/it'
j. tiklasola	'you (sg.) love him/her/it'
k. kilasola	'he/she/it loves him/her/it'
l. tiklasolan	'we love him/her/it'
m. nankiçlasolan	'you (pl.) love him/her/it'
n. kilasolan	'they love him/her/it'
o. tite:člasola	'you (sg.) love us'
p. te:člasola	'he/she/it loves us'
q. nante:člasolan	'they love us'
r. nam:ečlasola	'I love you (pl.)'
s. me:člasola	'he/she/it loves you (pl.)'
t. tame:člasolan	'we love you (pl.)'
u. me:člasolan	'they love you (pl.)'
v. nikinlasola	'I love them'
w. tikinlasola	'you (sg.) love them'
x. kinlasola	'he/she/it loves them'
y. tikinlasolan	'we love them'
z. nankinlasolan	'you (pl.) love them'
aa kinlasolan	'they love them'

7 Typology

CHAPTER OUTLINE

KEY TERMS

isolating
agglutinative
fusional
polysynthetic
index of synthesis
index of fusion
index of exponence
head-marking
dependent-marking
implicational
universal

In this chapter you will learn about morphological typology: how morphologists characterize the morphological systems of languages.

- ◆ We will begin by describing the morphological systems of five very different languages, looking at the kinds of lexeme formation and inflection that they display.
- ◆ Then we will discuss both traditional ways of classifying the morphology of languages and more contemporary ways of doing so.
- ◆ Finally, we will look at how both the family a language belongs to and the geographic area in which it is spoken can influence its typological classification.

7.1 Introduction

In this book, we have focused so far on formal processes of morphology that occur in many languages of the world; we have provided a sort of toolkit for forming new words from which languages can pick and choose. What we haven't looked at yet is what we could call 'the big picture': how does word formation work overall in specific languages and how can the morphological systems of particular languages vary from one another? In other words, rather than looking at specific processes and how they work, we can look at how languages exploit different parts of our toolkit to constitute their own unique systems of morphology. We can try to characterize the morphological systems of languages according to the sorts of morphological processes that they exploit. We can compare languages to each other to see if characteristics of their morphological systems correlate in any way with other parts of their grammars, say their syntax or their phonology. And we can look at the distribution of morphological patterns in terms of genetic relationships (language families) or areal tendencies (where languages are spoken in the world). Studying language from this sort of global perspective is the subject of linguistic **typology**.

7.2 Universals and Particulars: A Bit of Linguistic History

The history of linguistics has for centuries seen a tug of war between theories that emphasize universals – those things that are common to all human languages, perhaps because they are part of our common biological endowment – and particulars – those things that look unique and appear to distinguish languages from one another.

At the turn of the twentieth century, partly as a legacy of colonialism, linguists started studying indigenous languages of Africa, Asia, and the Americas more seriously. In North America, the tradition of American Structuralism stressed the uniqueness of languages, not surprising, considering the linguistic diversity of indigenous North American languages and their prodigious differences from one another and from better studied Indo-European languages. With the advent of Generative Grammar in the middle of the twentieth century, the pendulum began to swing in the other direction. Chomskians stressed what's universal in languages, and searched for ways to explain linguistic differences as the result of small choices that languages make from a universal set of options that our biological make-up, our hard-wiring for language as it were, makes available. All linguists today, whatever their theoretical commitments, recognize the importance of looking at a broad range of languages to observe both commonalities and differences. Understanding this universal set of options is ever more important today, with renewed efforts among linguists to study the many languages that are endangered.

Universals and particulars are both important: until we have a sense of the full range of particulars, we can only begin to confront the issue of universals. That's why studying the widest range of languages possible is so critical.

Do we know anything about morphological universals? Yes – and you've gotten a taste of what we know here. We know, for example, that there is a range of word formation strategies that appear in the languages of the world. And there are some conceivable sorts of word formation strategies that never occur. We know, for example, that there's no language so far that forms one sort of word from another – say nouns from verbs or verbs from nouns – by reversing the sounds of the words, or by infixing [p] after every third sound. But there are a lot of things we don't know – what are possible forms of reduplication or infixing, for example, and what is impossible. So the search for particulars and universals goes on in tandem.

7.3 The Genius of Languages: What's in Your Toolkit?

Students of linguistics often have a sense of excitement when they come upon data from languages they've never heard of before and discover how very different languages can look from one another. Although linguists in the generative tradition are always quick to stress that languages are more alike underlyingly than they seem superficially, what often strikes students first are the wonderfully exuberant ways in which languages can do things differently. Some of what gives this impression of difference is the unique way in which the morphology of languages can package different concepts in different forms.

The linguist Edward Sapir, writing early in the twentieth century, had a rather romantic name for the unique combination of processes that characterize the grammar of each language – he called it the “genius” of the language (Sapir 1921: 120):

For it must be obvious to anyone who has thought about the question at all or who has felt something of the spirit of a foreign language that there is such a thing as a basic plan, a certain cut, to each language. This type or plan or structural ‘genius’ of the language is something much more fundamental, much more pervasive than any single feature of it that we can mention, nor can we gain an adequate idea of its nature by a mere recital of the sundry facts that make up the grammar of the language.

These days, linguists might find quaint Sapir's idea that each language is imbued with something like a special spirit or soul that embodies its ‘basic plan’. But Sapir's idea of ‘genius’ comes close to that feeling that students have that the new languages they encounter are in some sense new creatures.

In this section we take a brief look at five very different languages – Turkish, Mandarin Chinese, Samoan, Latin, and Nishnaabemwin – to try to see something of this unique combination of morphological processes that constitutes at least one part of the genius of each language. All of these languages use morphology in one way or another, but each makes different choices from the universal toolbox of rule types that we have surveyed so far in this book. Some use predominantly one strategy, others many; some have lots of inflection, others almost none. But each has its own unique pattern.

7.3.1 Turkish (Turkic)

Imagine one word that means ‘were you one of those whom we are not going to be able to turn into Czechoslovakians?’ This may seem highly unlikely, but it’s possible in Turkish: the word is *çekoslovakyalılaştıramayacaklarımızdan mıydınız*, and it’s possible because Turkish is a language that delights in suffixation:

- (1) Turkish (Inkelas and Orgun 1998: 368)¹
- | | | | | | | | |
|------------------|----------|----------|---------|----------|---------|-------|--------|
| çekoslovakya - | lı - | laş - | tır - | ama - | yacak - | lar - | ımız - |
| Czechoslovakia - | from - | become - | CAUSE - | unable - | FUT - | PL | 1pl - |
| | | | | | | | |
| dan | mı - | ydı - | nız | | | | |
| ABL - | INTERR - | PAST - | 2PL | | | | |

Let’s look in detail at the pieces that make up this word. Note first that Turkish has a phonological rule called ‘vowel harmony’ which makes the vowels of suffixes agree with the preceding vowels in the base in backness and sometimes roundness (we’ll look more closely at this rule in [Chapter 9](#)). The first suffix that we encounter after the base is a suffix *-li*, which attaches to nouns to make personal nouns. With vowel harmony, the suffix *-li* will show up as *-lu* after the front round vowel *ö*:

- (2) *-li (-lu) personal nouns* (Lewis 1967: 60)
- | | | | |
|-------|-----------|----------|----------------|
| şehir | ‘city’ | şehir-li | ‘city dweller’ |
| köy | ‘village’ | köy-lu | ‘villager’ |

The next suffix *-laş* forms intransitive verbs from adjectives. Again, taking vowel harmony into account, the suffix can appear as *-leş* if it is preceded by a base with front vowels, and *-laş* if preceded by back vowels:

- (3) *-laş (-leş) intransitive verbs* (Lewis 1967: 228–9)
- | | | | |
|---------|-------------|-------------|----------------------|
| ölmez | ‘immortal’ | ölmez-leş | ‘become immortal’ |
| garp-lı | ‘West-from’ | garp-lı-laş | ‘become Westernized’ |

Next we have the suffix *-dir*, which forms causative verbs from intransitive verbs. Note that the [d] in the suffix shows up as the corresponding voiceless stop [t] if the preceding consonant is voiceless:

1. Inkelas and Orgun (1998) give this word in phonetic transcription. Here it is rewritten in Turkish orthography, as nothing hinges on the pronunciation.

- (4) *-dir (-dur, -dür, -tir, etc.) causative* (Lewis 1967: 144–5)
 don ‘freeze’ don-dur ‘cause to freeze’
 öl ‘die’ öl-dür ‘kill’ (= ‘cause to die’)

The following suffix is called the ‘impotential’, which essentially means ‘not able’:

- (5) *-eme (-ama) impotential* (Lewis 1967: 151)
 gel-eme ‘unable to come’
 anlı-ama ‘unable to understand’

Next is the future suffix *-ecek*:

- (6) *-ecek (-acak) future* (Lewis 1967: 114)
 bul-acak ‘will find’
 tanı-acak ‘will recognize’

Then the plural suffix *-ler*:

- (7) *-ler (-lar) plural* (Lewis 1967: 29)
 kız ‘girl’ kız-lar ‘girls’
 el ‘hand’ el-ler ‘hands’

And then, we get the first person possessive suffix *-ımız*, the ablative marker *-dan* (which essentially means ‘from’), the question marker *-mı*, the locative marker *-dı* (which means ‘at’), and the second person plural verb marker *-mız*. As you can see, each of these suffixes can occur in simpler words, but they can all be used together to form the enormously long and complex word we started with. If we put it all back together again, we get something like (8):

- | | |
|--|--|
| (8) çekoslovakya-lı | ‘someone from Czechoslovakia’ =
‘Czechoslovakian’ |
| çekoslovakya-lı-laş | ‘become Czech’ |
| çekoslovakya-lı-laş-tır | ‘cause to become Czech’ |
| çekoslovakya-lı-laş-tır-ama | ‘unable to cause to become Czech’ |
| çekoslovakya-lı-laş-tır-ama-
(y)-acak | ‘will be unable to cause to become
Czech’ |
| çekoslovakya-lı-laş-tır-ama-
(y)-acak-lar | ‘will be unable to cause to become
Czech-pl.’ |
| çekoslovakya-lı-laş-tır-ama-
(y)-acak-lar-ımız | ‘we will be unable to cause to
become Czech-pl.’ |
| çekoslovakya-lı-laş-tır-ama-
(y)-acak-lar-ımız-dan | ‘from (those) we will be unable to
cause to become Czech-pl.’ |
| çekoslovakya-lı-laş-tır-ama-
(y)-acak-lar-ımız-dan-mı-
ydı-nız | ‘were you from (those) we will
be unable to cause to become
Czech-pl.’ |

Although the predominant way of forming words in Turkish is through suffixation, it also has a process of compounding, as the examples in (9) show:

(9) *Turkish compounds* (Lewis 1967: 231–3)

Noun + noun	baba +	anne	babaaanne
	father	mother	paternal grandmother
Adjective + noun	baş +	bakan	başbakan
	head	minister	prime minister
	kırk +	ayak	kırkayak
	forty	foot	centipede
	büyük +	anne	büyükanne
	great	mother	grandmother

As the examples in (9) show, Turkish compounds are right-headed.

We can see from our lengthy discussion of the example in (1) that Turkish uses suffixation for both derivation and inflection. In addition to marking number on nouns, Turkish also marks case, as the paradigm in (10) shows:

(10)	ev	‘house’
	evi	Definite-accusative
	evin	Genitive
	eve	Dative
	evde	Locative
	evden	Ablative

Turkish verbs are inflected for person and number, and can appear in a number of different tenses, including present, past, future, and conditional. There are affixes to make verbs negative or interrogative, and verbs can mark other distinctions as well. All of these inflections are suffixes; verb forms can be quite long and complex.

What this brief description of Turkish morphology shows is that a language can have wildly abundant morphology, and yet use no more than a couple of tools from our universal toolkit. Turkish is an overwhelmingly, exuberantly suffixing language, using suffixes for both lexeme formation and inflection, but it has no processes of prefixation to speak of. What few prefixes can be found in Turkish are always on borrowed words, and essentially are not part of the native system of word formation of Turkish.

7.3.2 Mandarin Chinese (Sino-Tibetan)

Where the genius of Turkish morphology lies in the exuberance of its processes of suffixation, Mandarin Chinese makes entirely different choices from our universal toolkit, although it too concentrates on just a few tools.

According to Li and Thompson (1971), Mandarin has no processes of prefixation to speak of, and it has only a tiny handful of suffixes, among them the three in (11):²

2. Li and Thompson give a few examples of prefixes, but it's not clear why they treat them as prefixes rather than as the first elements of compounds.

(11) *Li and Thompson (1971: 41–3)*

-xué	dòngwù-xué	‘animal-ology’ = ‘zoology’
	shèhuì-xué	‘society-ology’ = ‘sociology’
	zhé-xué	‘philosophy-ology’ = ‘study of philosophy’
-jiā	kēxué-jīa	‘science-ist’ = ‘scientist’
	yùndòng-jīa	‘athletics-ist’ = ‘athlete’
	zhèngzhì-jīa	‘politics-ist’ = ‘politician’
-huà	gōngyè-huà	‘industry-ize’ = ‘industrialize’
	tóng-huà	‘similar-ize’ = ‘assimilate’

The suffix *-xué* attaches to nouns to make nouns meaning ‘the study of X’, and *-jiā* makes personal nouns, also from other nouns. Notice that in the example *kēxué-jīa*, both suffixes occur. The third suffix *-huà* makes verbs from nouns and adjectives. Suffixation is quite limited in Mandarin though: there are relatively few suffixes, and we certainly do not find words of the complexity of those in Turkish or even of words in English.

Mandarin also has full reduplication, which it uses for two purposes. Verbs can be reduplicated to form derived verbs meaning ‘X a little’:

(12) *Li and Thompson (1971: 29)*

jiāo	‘teach’	jiāo-jīao	‘teach a little’
shuō	‘say’	shuō-shuō	‘say a little’
xiē	‘rest’	xiē-xiē	‘rest a little’

And some adjectives can be reduplicated to make intensive adjectives, that is, adjectives that mean ‘very X’:

(13) *Li and Thompson (1971: 33)*

pàng	‘fat’	pàng-pàng	‘very fat’
hóng	‘red’	hóng-hóng	‘very red’
yuán	‘round’	yuán-yuán	‘very round’

Though Mandarin is relatively poor in affixation and reduplication, it is incredibly rich in compounding. Mandarin has not only compound nouns and compound adjectives, as English does, but also all sorts of compound verbs, as the examples in (14) show:

(14) *Examples from Li and Thompson (1971: 49–55), Ceccagno (undated ms. 3–8), and Li (1995: 256)*

- a. [N + N]_N

hè-mǎ	‘river-horse’ = ‘hippopotamus’
hǎi-gǒu	‘sea-dog’ = ‘seal’
chún-gāo	‘lip-ointment’ = ‘lipstick’
- b. [N + N]_N shū-guǒ ‘vegetable-fruit’ = ‘vegetables and fruit’
- c. [A + A]_A liàng-lì ‘bright-beautiful’ ‘bright and beautiful’
- d. [A + N]_N

hēi-chē	‘black-vehicle’ = ‘illegal vehicle’
zhǔ-yè	‘main-page’ = ‘home page’
- e. [V + N]_N

jiān-shì	‘supervise-matter’ = ‘supervisor’
wén-xiōng	‘hide-breast’ = ‘bra’

f. [V + N] _v	dài-gǎng	‘wait for-post’ = ‘wait for a job’
	jìn-dú	‘prohibit-poison’ = ‘ban (sale/use of) drugs’
g. [A + V] _v	gōng-shì	‘public-show’ = ‘make public’
h. [V + V] _v	dǎ-pò	‘hit-broken’
	lā-kāi	‘pull-open’
	zhui-lei	‘chase-tired’

As the examples in (14) show, Mandarin has many different types of compound, indeed, many more types than English, which as we’ve seen has very productive compounding processes. In Mandarin, some compounds are attributive, for example those in (14a) and (14d). The examples in (14b, c) are coordinative compounds. And some are subordinative, for example those in (14e, f, g). Most of these compounds are endocentric (14a, b, c, d, f, g, h), but those in (14e) are exocentric. Some are right-headed (14a, c, d, f, g), some have two heads (the coordinative compounds in (14b, c)), and some are left-headed (14h).

The examples we have given above all concern lexeme formation in Mandarin because Mandarin has rather little in the way of inflection. Nouns are not inflected for number, nor do verbs inflect for person or number. Tense is not marked morphologically and aspect is marked only by separate particles that appear after the verb. Mandarin does have a system of noun classifiers that are used when counting or otherwise quantifying nouns, but again separate particles rather than affixes are used to mark the class of particular nouns. So we can return to the sentences we looked at in [Chapter 1](#) and see them in the wider context of Mandarin morphology:

(15) Wo jian guo yi zhi chang jing lu.
I see EXP³ one CLASSIFIER giraffe

(16) Wo jian guo liang zhi chang jing lu
I see EXP two CLASSIFIER giraffe

The verb *jian* ‘see’ doesn’t change its form from one aspect to another, nor does the noun *chang jing lu* ‘giraffe’ change from singular to plural. In order to signal more than one giraffe, one needs to signify a specific number or quantity.

What this shows us is that Mandarin is just as able as Turkish to come up with new words and to express grammatical distinctions, but its strategy for doing so makes use of different means.

7.3.3 Samoan (Austronesian)

What makes Samoan an interesting contrast to Turkish and Mandarin is that it uses a wide variety of word formation processes without seeming to favor one over another. To pursue our mechanical metaphor, its

3. In [Chapter 1](#) we translated *guo* as ‘past’, but this was not quite right. The morpheme *guo* marks what Li and Thompson (1971: 226) call “experiential aspect,” which signals that the speaker has experienced the event. It is, however, often used in the context of an event that has already happened.

toolbag is chock-full of different tools. In this language we can find prefixation, suffixation, and circumfixation, both partial and full reduplication, and also to some extent compounding. There's also even a bit of internal stem change in the form of a morphological process of vowel lengthening. Here are some examples:

- (17) *Prefixation: fa'a 'causative'* (Mosel and Hovdhaugen 1992: 175–6)

alu	'go'	fa'aalu	'make go'
goto	'sink'	fa'agoto	'make sink'
ga'o	'fat'	fa'aga'o	'apply grease to'
māsima	'salt'	fa'amāsima	'salt' = 'put salt on'

The prefix *fa'a* can be put on either verbs or nouns to make verbs meaning 'cause X' or 'make X' or 'put X on'.

- (18) *Circumfixation: fe-a'i 'reciprocal'* (Mosel and Hovdhaugen 1992: 182)

finau	'quarrel'	fefinaua'i	'quarrel with one another'
logo	'inform'	felogoa'i	'consult with one another'
mata	'look'	femātaa'i	'look at one another'

Although Mosel and Hovdhaugen say that prefixes are usually mutually exclusive – that is, there can only be one in a word – the circumfix *fe-a'i* can occur outside the prefix *fa'a*, as you see in the word in (19):

- (19) fe-fa'a-māfanafaa-a'i
 recip-cause-warm- recip
 'be of comfort to one another'

In addition to prefixes and circumfixes, Samoan can also form words by suffixation:

- (20) *Suffixation: -ga forms abstract nouns from verbs* (Mosel and Hovdhaugen 1992: 195)

amo	'carry'	amoga	'carrying'
a'o	'learn'	a'oga	'education'
savali	'walk'	savaliga	'a walk'

Suffixing *-ga* to a verb and lengthening the first vowel of the verb stem forms another kind of derived noun which can be concrete and often, according to Mosel and Hovdhaugen (1992: 195), has a flavor of plurality:

- (21) *Suffixation of -ga and vowel lengthening*
- | | | | |
|--------|---------|----------|-------------------------------------|
| amo | 'carry' | āmoga | 'person(s) carrying loads on yokes' |
| a'o | 'learn' | ā'oga | 'school' |
| savali | 'walk' | sāvaliga | 'people on march' |

The vowel lengthening that occurs in this process can be considered a form of internal stem change.

Samoan is also rich in processes of reduplication, as we already saw in Section 5.5. As we saw there, Samoan has a process of partial reduplication that forms verbs from nouns. To repeat example (16) from Chapter 5:

(22) *Partial reduplication: $N \rightarrow V$*

lafo	'plot of land'	lalafo	'clear land'
lago	'pillow, bolster'	lalago	'rest, keep steady'
pine	'pin, peg'	pipine	'secure with pegs'

Partial reduplication can also be used to make ergative verbs from nonergative verbs:

(23) *Partial reduplication: $V_{\text{nonergative}} \rightarrow V_{\text{ergative}}$ (Mosel and Hovdhaugen 1992: 222–3)*

lo'u	'bent'	lolo'u	'bend'
motu	'break (nonerg.)'	momotu	'break (erg.)'
sa'e	'overturn'	sasa'e	'cause to capsiz'

In both cases, partial reduplication copies the first consonant and vowel of the base.

New words are also formed in Samoan by full reduplication. We saw one example (15b) in Section 5.5, which is repeated in (24), and another example is given in (25):

(24) *Full reduplication: $V \rightarrow N$ (Mosel and Hovdhaugen 1992: 229)*

'apa	'beat, lash'	'apa'apa	'wing, fin'
au	'flow on, roll on'	auau	'current'
solo	'wipe, dry'	solosolo	'handkerchief'

(25) *Full reduplication: frequentative/intensive*

a'a	'kick'	a'aa'a	'kick repeatedly'
'emo	'blink, flash'	'emo'emo	'twinkle'
fo'i	'return'	fo'ifo'i	'keep going back'

In (25) we see a process of full reduplication that takes verbs and makes them into **frequentatives**, that is, forms that mean 'X repeatedly', or **intensives**, forms that mean 'X a lot'.

Finally, Samoan also has compounding, as the examples in (26) show:

(26) *Compounding (Mosel and Hovdhaugen 1992: 241–3)*

alagā	'source'	'oa	'valuable goods'	alagā'oa	'source of wealth'
'aliti	'bed'	tai	'sea'	'aliti tai	'seabed'
fāsi	'piece of'	moli	'soap'	fāsimoli	'piece of soap'
suā	'liquid'	esi	'paw paw'	suāesi	'paw paw soup'

These examples are all left-headed endocentric attributive compounds. As we can see, although compounding in Samoan is possible, this language has nowhere near the richness of compound types that can be found in Mandarin.

Interestingly, although Samoan sentences express case relations (ergative/absolutive) and clauses are marked for tense, aspect, and mood, Samoan has no inflectional paradigms (Mosel and Hovdhaugen 1992: 169). In fact, relations like case, tense, aspect, and mood are expressed by independent particles, rather than by prefixes, suffixes, or reduplication, in this language; hence most of our examples here have been derivational.

7.3.4 Latin (Indo-European)

Like Turkish, and unlike Mandarin and Samoan, Latin is a heavily inflected language. And like Turkish, its inflections are almost entirely suffixal.⁴ However, its inflection looks rather different from Turkish inflection in that often several meanings are combined into a single inflectional morpheme in Latin. Its toolbox is somewhat larger for derivation than for inflection, with some prefixation and compounding in addition to suffixation. Indeed, you will probably recognize elements of Latin derivational morphology, as many of them have been borrowed into English. We will look at inflection first, then derivation.

Latin nouns are inflected for case, number, and gender, and adjectives are inflected to agree with them. (27) shows the paradigm for the feminine noun *puella* ‘girl’, and (28) a noun phrase with an agreeing adjective:

(27)	‘girl’	<i>Singular</i>	<i>Plural</i>
	Nom.	<i>puella</i>	<i>puellae</i>
	Gen.	<i>puellae</i>	<i>puellārum</i>
	Dat.	<i>puellae</i>	<i>puellīs</i>
	Acc.	<i>puellam</i>	<i>puellās</i>
	Abl.	<i>puellā</i>	<i>puellīs</i>

(28)	<i>bona puella</i>	good-NOM.SG	girl-NOM.SG	‘good girl’
	<i>bonae puellae</i>	good-NOM.PL	girls-NOM.PL	‘good girls’

Each inflection carries a combination of meanings that includes case, number, and gender. For example, the morpheme *-ārum* is used in the genitive plural, and in addition, signals that this noun belongs to the first Latin declension, almost all of whose members are feminine in gender.

Verbs have a number of different stems which form the basis of inflectional paradigms that show aspect (imperfect vs. perfect) and voice (active vs. passive), as well as person and number. A portion of the paradigms for the verbs ‘love’ and ‘warn’ are shown in (29) (these are the same examples you looked at in exercise 6 of [Chapter 6](#)):

(29)	<i>amābō</i>	‘I will love’	<i>monēbō</i>	‘I will warn’
	<i>amābis</i>	‘you will love’	<i>monēbis</i>	‘you will warn’
	<i>amābit</i>	‘he/she will love’	<i>monēbit</i>	‘he/she will warn’
	<i>amābimus</i>	‘we will love’	<i>monēbimus</i>	‘we will warn’
	<i>amābitis</i>	‘you (pl.) will love’	<i>monēbitis</i>	‘you (pl.) will warn’
	<i>amābunt</i>	‘they will love’	<i>monēbunt</i>	‘they will warn’
	<i>amāvī</i>	‘I loved’	<i>monuī</i>	‘I warned’
	<i>amāvistī</i>	‘you loved’	<i>monuistī</i>	‘you warned’
	<i>amāvit</i>	‘he/she loved’	<i>monuit</i>	‘he/she warned’
	<i>amāvimus</i>	‘we loved’	<i>monuimus</i>	‘we warned’
	<i>amāvistis</i>	‘you (pl.) loved’	<i>monuistis</i>	‘you (pl.) warned’
	<i>amāvērunt</i>	‘they loved’	<i>monuērunt</i>	‘they warned’

4. Latin has a small amount of reduplication and infixation in its verbal paradigms, but both processes occur only in a small set of verbs, so we will not go into them here.

These verb forms are built on one of the stem forms, called the Theme Vowel stem (*amā-*, *monē-*) to which a future suffix *-bi* or a perfect suffix *-v* is attached. Then person and number suffixes are attached. Interestingly, different person and number affixes are used in the past than in other tenses:

(30) *Person and number suffixes in Latin verbs*

Non-past

singular	1	-ō	plural	1	-imus
	2	-s		2	-itis
	3	-t		3	-unt

Past

singular	1	-ī	plural	1	-imus
	2	-istī		2	-istis
	3	-it		3	-ērunt

In the non-past, the suffixes combine person and number; in some sense, however, the second set of suffixes also signal past tense in addition to person and number, since they are only used in the past tense.

Latin has both derivational suffixes and prefixes. For example, it forms abstract nouns from verb roots by adding the suffix *-or*:

- (31) timor ‘fear’ timēre ‘to fear’
 amor ‘love’ amāre ‘to love’

The suffix *-men* attaches to either roots or theme vowel stems to form nouns that denote the result of an action:

- (32) agmen ‘line of march, band’ agere ‘to lead’
 certāmen ‘contest, battle’ certāre ‘to contend’

The prefix *amb-* attaches to verbs and means ‘around’ and the prefix *in-* attaches to adjectives to form negative adjectives:

- (33) īre ‘to go’ amb-īre ‘to go about’
 sānus ‘sane’ in-sānus ‘insane’

Latin does not use compounding as much as English and Germanic languages do, but it does have some compounds:

- (34) flōs, flōris ‘flower’ coma ‘hair’ flōri-comus ‘flower-crowned’
 āla ‘wing’ pēs ‘foot’ āli-pēs ‘wing-footed’

When two roots are put together into a compound, the linking vowel *-i* is used between them.

7.3.5 Nishnaabemwin (Algic)

The final language we will look at is Nishnaabemwin, an Algic language spoken in southern Ontario, Canada (Valentine 2001). It makes heavy use of affixation, especially suffixation, and has an extremely rich system of inflection. What’s most interesting about Nishnaabemwin, however, is the way that it combines bound morphemes to form new lexemes.

Let's look first at inflection. Nouns have either animate or inanimate gender. They can be inflected for number, and have different forms for diminutive, pejorative, and what is called 'contemptive', a suffix that adds a negative meaning, but one that is less strongly negative than the pejorative suffix. Nouns can also occur in a locative form, and what is called the 'obviative' form, which serves to distance a noun from the speaker in a narrative. Finally, there are prefixes and suffixes that indicate possession of a noun. Some of the relevant forms for the noun *zhiishiib* 'duck' are given in (35):

(35)	<i>zhiishiib</i>	'duck'
	<i>zhiishiib-ag</i>	Plural
	<i>zhiishiib-an</i>	Obviative
	<i>zhiishiib-enh</i>	Contemptive ⁵
	<i>zhiishiib-ens</i>	Diminutive
	<i>zhiishiib-ish</i>	Pejorative
	<i>zhiishiib-ing</i>	Locative
	<i>n-zhiishiib-im</i>	'my duck'

These inflections are not mutually exclusive. If they occur in combination, the diminutive, for example, must precede the pejorative, which in turn precedes the suffix that occurs in the possessive. Nor are these the only inflections that can appear on nouns: this is just a small selection!

Verb inflection is even more complex than noun inflection, and we cannot really do justice to it here. But just to give you a taste of how very intricate the inflection of verbs is in Nishnaabemwin, consider Valentine's (2001: 219) description of the inflectional possibilities of intransitive verbs that take animate subjects:

For the INDEPENDENT and CONJUNCT ORDERS, there are theoretically nine person/number/obviation categories, four MODES (and ITERATIVES in the conjunct), two POLARITIES, three TENSES, and two functions (VERBAL and PARTICIPIAL in the conjunct) creating 882 inflectional combinations. For the IMPERATIVE ORDER, there are three person/number combinations, and three modes, creating theoretically nine inflectional combinations, making a total of roughly 890 forms for the animate intransitive verb.

The independent form of the verb is used in main clauses, and differently inflected forms are used in subordinate clauses (the conjunct form) and in imperatives. By polarities, Valentine means that verbs have different forms for positive and negative. Compare this to the four forms of the verb *walk* in English (*walk, walks, walked, walking*); English uses the same forms of the verb in main and subordinate clauses, and uses a separate word for negation.

Derivation is no less complex than inflection in this language: words rarely consist of a single root morpheme. Instead, various bound

5. I have constructed this form and the next one. Valentine (2001) does not cite these forms himself.

morphemes are joined together to form words. Intransitive verbs, for example, frequently consist of two or three pieces. The last piece, called the ‘final’, expresses a verbal concept. The first piece, called the ‘initial’, expresses something that modifies the verbal concept; in English we would express similar concepts with separate adjectives, adverbs, or prepositions:

- (36) Examples from *Valentine* (2001: 327) with initial giin- 'sharp'
 giin'zi 'be sharp (of tongue)'-ANIM. SUBJ.
 giinaa 'be sharp'-INANIM. SUBJ.

Verbs may also have a third piece that occurs between the initial and the final; this is called the ‘medial’ and corresponds to what in English would usually be a nominal concept (Valentine 2001: 330, 334):

- | | | |
|------|-------------------------------------|---|
| (37) | dewnike 'have an ache in one's arm' | dew 'sore' + nik 'arm' + e
'have' |
| | bookjaane 'have a broken nose' | book 'broken' + jaan
'nose' + e 'have' |
| | gaagijndbe 'have a sore head' | gaagij 'sore' + ndib 'head'
+ e 'have' |

Such forms can undergo further derivation by adding another final element, so from *gaagijjndbe* 'have a sore head' we can get *gaagijjnd-bekaazo*, which means 'pretend to have a sore head' by adding the final *-kaazo*, which means 'pretend to'. And of course each of these verbs can then take various inflectional elements.

Nouns can be made up of several bound morphemes as well. For example, the nouns in (38a) are made up of an initial bound element that has an adjectival sort of meaning, and a second bound element that means 'thing'. Those in (38b) have an initial element that is a free form, and a bound final element that means 'building, habitation':

- (38) a. From *Valentine* (2001: 484)
getehii 'old thing' gete-'old' -hii 'thing'
shkihii 'new thing' oshk-'new' -hii 'thing'
- b. bzhikiiwgamig 'cowshed' bizhikiw 'cow' -gamigw 'building'
gookooshgamig 'pig sty' gookoosh 'pig' -gamigw 'building'

Valentine uses terms like 'initial', 'medial', and 'final', rather than calling the bound morphemes that make up these forms 'prefixes' or 'suffixes', probably because these morphemes are much more like roots than like affixes in their meanings. We might therefore think of them as bound bases.

There are some derivational suffixes in Nishnaabemwin, however. For example the suffix *-aagan* creates nouns from one class of transitive verbs:

- (39) naabkawaagan ‘scarf, necklace’ naabkaw ‘wear ANIM.NOUN around
neck’
noodaagan ‘employee’ noodaw ‘hire ANIM.NOUN’

And Nishnaabemwin also has compounds, which differ from the forms in (38) by being composed of two independent stems (Valentine 2001: 516–17):

- | | | | |
|------|------------------|---------------|-------------------|
| (40) | jiibaakwe-kik | ‘cooking pot’ | (cook + pot) |
| | shkode-daaban | ‘train’ | (fire + vehicle) |
| | waasgamg-kosmaan | ‘bell pepper’ | (pepper + squash) |

One thing that is striking about the morphological system of Nishnaabemwin is the overall complexity of words: words rarely consist of a single morpheme, and frequently consist of many morphemes.

7.3.6 Summary

Comparing these languages, we can see a bit of what Sapir means about the ‘genius’ of a language. Although we don’t need to romanticize the unique character of each language, studying morphology opens our eyes to the different mixture and balance of word formation processes to be found in individual languages. Each language has a different combination of word formation processes that gives the language its unique character. But we should keep in mind as we wonder at all this diversity that we should always be on the lookout for the commonalities or universals that mark all these languages as human languages.

Challenge

Find a grammar of a language you know nothing about. Write a two- or three-page description of the sorts of morphology that your chosen language displays, thinking about both inflection and derivation, and about the different formal processes of word formation rules your language displays. We will look at these again shortly.

7.4 Ways of Characterizing Languages

Up to this point we have viewed the morphological systems of languages in an impressionistic way, looking at the combination of inflectional and derivational processes that give the language its overall morphological pattern. Morphologists continually seek to go beyond simple descriptions, however. In looking at the morphological systems of individual languages, we are always looking for patterns. We are interested in what sorts of morphological rules we might expect to find in languages and what they tell us about the general faculty of human language. We are also interested in classifying languages by looking at whether particular sorts of traits cluster together, and whether those clusters tell us something deeper about the nature of language. We will return to the nature of morphological rules in Chapter 10. Here, we will look more closely at different ways we may classify the morphology of

languages, and how various classifications illuminate the nature of language. This latter enterprise is called typology. We start this section by looking at a very traditional way of classifying languages, and then look at more contemporary schemes of morphological typology.

WALS: The World Atlas of Linguistic Structures

A great resource for student linguists is the World Atlas of Linguistic Structures, or WALS, which can be found online at <https://wals.info>. WALS is a website that provides information on all sorts of phonological, syntactic, and morphological characteristics of the languages of the world, surveying everything from phonological inventories to types of clauses and illustrating the cross-linguistic results in easily accessible maps. Each of its 150 or so chapters covers a particular characteristic of language and provides an explanation of that phenomenon along with examples. For morphology, we can find lots of information on inflection, especially on the expression of person, number, gender, case, tense, aspect, and number in the languages of the world.

For example, Chapter 26 surveys the extent to which inflectional morphology is expressed by prefixes or suffixes (or by both or neither). In the vast majority of languages that exhibit inflection, it is expressed by suffixes rather than prefixes. However, there is a concentration of strongly prefixing languages in sub-Saharan Africa and a cluster of languages that use neither prefixes nor suffixes for inflection in Southeast Asia (these languages display little or no inflection at all), but even a quick glance at the map accompanying Chapter 26 shows the prevalence of suffixing.

Chapter 27 looks at the extent to which reduplication – either partial or full – can be found in the languages of the world. Of the 368 languages surveyed for this chapter, 278 of them are said to have both full and partial reduplication, 35 to have only full reduplication, and 55 to have no productive reduplication at all. English is one of this last group, so WALS allows us to see one way in which our language is quite atypical!

7.4.1 The Fourfold Classification

Morphological typology has a long history, going back at least to the early nineteenth century in the work of [Wilhelm von Humboldt \(1836; reference in Comrie 1981/1989\)](#). In this tradition, also developed by the linguist Edward Sapir, whom we mentioned above, it was common to divide languages into four morphological types: isolating (or analytic), agglutinative, fusional, and polysynthetic.

An **isolating** or **analytic** language is one in which each word consists of one and only one morpheme. Vietnamese is often cited as an example of an isolating language. For example, nouns do not inflect for plurality.

The noun *đồng hồ* means ‘watch’ or ‘watches’. If one wants to be specific about how many watches are in question, it is possible to use a numeral and then a noun classifier before the noun (Nguyen and Jorden 1969: 119), as (41a) shows:

- (41) a. hai cái đồng hồ
 two CL watch
 ‘two watches’
 b. Mai tôi làm cái đó Tôi làm cái đó hôm qua
 tomorrow I do CL that I do CL that yesterday
 ‘I will do that tomorrow’ ‘I did that yesterday’

Similarly, as (41b) illustrates, verbs do not inflect for tense in Vietnamese. Instead, if one wants to be specific about the time of an event, it is necessary to use specific adverbs like ‘tomorrow’ or ‘yesterday’.

Of the languages we have profiled in Section 7.3, Mandarin comes closest to being an isolating language. Although Mandarin has abundant compounding, it has little that would count as morphological inflection. Like Vietnamese, plurality, tense, and aspect are all expressed by separate words.

Unlike isolating languages, **agglutinative** languages have complex words. Furthermore, those words are easily segmented into separate morphemes and each morpheme carries a single chunk of meaning. In our grammatical sketches in 7.3, Turkish is the language that comes closest to an agglutinative ideal. For example, we gave the various case forms of the Turkish noun ‘house’ in (10), repeated here (42):

- (42) ev ‘house’
 evi Definite-accusative
 evin Genitive
 eve Dative
 evde Locative
 evden Ablative

To form the plurals of these nouns, all one needs to do is add the morpheme *-ler*, which goes after the root and before the case endings:

- (43) evler ‘house-plural’
 evleri Plural definite-accusative
 evlerin Plural genitive
 evlere Plural dative
 evlerde Plural locative
 evlerden Plural ablative

The two sorts of morphemes are easily separated in terms of both form and meaning.

A **fusional** language, like an agglutinative language, allows complex words, but its morphemes are not necessarily easily segmentable: several meanings may be packed into each morpheme, and sometimes it may be hard to decide where one morpheme ends and another one starts. Latin is a good example of a fusional language. We can, for

(44)	'girl'	<i>Singular</i>	<i>Plural</i>
	Nom.	puella	puellae
	Gen.	puellae	puellārum
	Dat.	puellae	puellīs
	Acc.	puellam	puellīs
	Abl.	puellā	puellīs

The final morphological type is called **polysynthetic**. In a polysynthetic language words are frequently extremely complex, consisting of many morphemes, some of which have meanings that are typically expressed by separate lexemes in other languages. In our grammatical sketches in **Section 7.3**, Nishnaabemwin is a language that could easily be characterized as polysynthetic. Remember that it not only has an intricate inflectional system, but forms complex words out of two or more bound bases. We repeat the examples from (37) here as an illustration of polysynthesis (**Valentine 2001**: 330, 334):

- | | | |
|------|--|---|
| (45) | dewnike 'have an ache
in one's arm' | dew 'sore' + nik 'arm' + e 'have' |
| | bookjaane 'have a broken nose' | book 'broken' + jaan 'nose' + e 'have' |
| | gaagiijndbe 'have a sore head' | gaagiij 'sore' + ndib 'head' + e 'have' |

7.4.2 The Indexes of Synthesis, Fusion, and Exponence

The problem with the traditional fourfold classification is that languages rarely fall neatly into one of the four classes. For example, English is not quite an isolating language – it has some inflection – but it is certainly not an agglutinating or a fusional language (Old English was much closer to being a fusional language, though). Another problem is that sometimes the inflectional system of a language falls into one category, but the derivational system fits better into another. English again can serve as an illustration: English derivational morphology is actually not that far from being agglutinating, as an example like *operationalizability* (*operat-ion-al-iz-able-ity*) suggests.

One way of dealing with these problems is to give up the fourfold classification in favor of three different scales which can be used to gauge different aspects of the overall morphological behavior of a language: the index of synthesis, the index of fusion, and the index of exponence.

The **index of synthesis** looks at how many morphemes there are per word in a language. Isolating languages will have few morphemes per word – in the most extreme cases, only one morpheme per word. Agglutinative or polysynthetic languages, however, will typically have many morphemes per word. And because this is a scale, languages like Samoan or English can fall somewhere in between the extremes.

The **index of fusion** is a measure of the degree to which morphemes are phonologically separable from their bases. In languages with a low degree of fusion, the phonological boundaries between morphemes are clear either because the languages are isolating in the classical sense or because they are agglutinative. In languages higher on the index of fusion, the phonological boundaries become increasingly less clear. Languages like Mandarin or Turkish would therefore be low on this index. Languages like Arabic or Hebrew with templatic (root and pattern) morphology would be high on the index of fusion, and languages like Latin would be somewhere in between.

The **index of exponence** measures how many meanings or inflectional features can be expressed simultaneously in a single morpheme in a language. We might expect, typically, for each morpheme to express a single meaning or feature – past tense, say, or plural. But in many languages a single morpheme can combine several meanings. Consider the example in (46) from the Chapacuran language Wari', spoken in Brazil (WALS, ch. 21; Everett and Kern 1997: 339):

- (46) Toc na com
 drink.SG 3SG.REAL.NONFUT.ACTIVE water
 'He is drinking water'

In the Wari' example, the first morpheme *toc* combines the meaning drink with singular. The second *na* combines third person with three other inflectional meanings. This language, then, allows what is known as **cumulative exponence**. Turkish, however, has a tendency to be low on the index of exponence, since typically each morpheme conveys only one meaning or inflectional feature. WALS uses the three indexes to look at inflectional morphology in the languages of the world. But there is no reason why we could not also look at the derivational and inflectional morphologies of a language separately and see where they fit on the scales. In terms of inflection, English would be low on the index of synthesis, but we might place it higher on that scale if we're looking at English derivation, since many words in the language are formed by compounding, prefixation, or suffixation. Similarly, we might class English higher on the index of exponence if we're looking at verbal inflection (the suffix *-s* carries the meanings 'third person', 'present', and 'singular' packed together in a form like *walks*) than if we're looking at derivation, where each morpheme typically has one distinct meaning.

Challenge

Take a look at the grammatical sketch of an unfamiliar language that you made earlier in this chapter. How would you characterize your language in terms of the fourfold classification? Where would you place it in terms of the index of synthesis, the index of exponence, and the index of fusion? Does your language pose any special problems for these means of classification?

7.4.3 Head- versus Dependent-Marking

Above, we have looked at the morphologies of languages in terms of the ease with which words can be segmented and the relationship between meaning and form in morphemes. There are other things we can look at, however, in classifying and comparing languages.

One thing we can look at is the way that morphology signals the relationship between words in phrases. The main element in each syntactic phrase is called its **head**; the head of a noun phrase (NP) is the noun, the head of a verb phrase (VP) is the verb, and so on. The other elements that combine with the head to become a phrase might be called the **dependents** of the head. Dependents of a noun can be adjectives, determiners, or possessives, and dependents of a verb can be its subject or object. Languages can choose to mark relationships between the head and its dependents in different ways: the relationship can be marked exclusively on the head, or exclusively on the dependent, or on both or neither. If the relationship is marked by some morpheme on the dependent, this is called **dependent-marking**, and if it is marked on the head, it is called **head-marking**.

As illustrated in (47a), the relationship between the head noun and its possessor is marked on the possessor in English, but on the head in Hungarian (47b) (examples from Nichols 1986: 57):

- (47) a. *English*
- | | |
|-----------|-------|
| the man's | house |
| dependent | head |
- b. *Hungarian (Uralic)*
- | | |
|-------------------|-------------|
| as | ember ház-a |
| the man | house-3sg |
| dependent | head |
| 'the man's house' | |

The NP in English therefore shows dependent-marking between a possessor and the head noun, whereas the NP in Hungarian shows head-marking.

Whole clauses may also exhibit either dependent-marking or head-marking. In Dyirbal (Pama-Nyungan), NPs are case-marked to show their relationship to the verb, which Nichols considers to be the head of the sentence:⁶

6. Not all linguists agree with her on this point, however. For example, in the generative tradition, Tense is considered to be the head of a clause.

(48) From *Nichols (1986: 61)*

balan	ɖugumbil	banul	yaŋangu	banɣu	yuguɲu	balgan
ART.NOM	woman.NOM	ART.ERG	man.ERG	ART.INSTR	stick.INS	hit
dependent		dependent		dependent		head

'The man is hitting the woman with a stick.'

Dyirbal would therefore be considered a dependent-marking language within clauses.

In contrast, Tzutujil (Mayan) shows head-marking within clauses: the verb is marked for the person and number of its subject and object, but there is no marking on the subject and object themselves to show their function in the clause (*Nichols 1986: 61*):

(49)	x-Ø-keetij	tzyaq	ch'ooyaa7
	ASP-3SG.-3PL.-ate	clothes	rats
	head	dependent	dependent

'Rats ate the clothes'

As I mentioned above, it is also possible for languages to show neither head-marking nor dependent-marking. For example, a language that is isolating and has no inflection would have neither head- nor dependent-marking. On the other hand, it is possible for a language to have inflectional markings on both the head and its dependents. Turkish, for example, marks both the possessor and the possessed noun in an NP:

(50)	ev-in	kapi-sı
	house-GEN	door-3SG
	dependent	head

'the door of the house'

When the relationship between the dependent and the head is marked on both constituents, we have what is called **double-marking**.

7.4.4 Correlations

In the last sections we have seen various ways in which the morphologies of languages can be classified. Typologists are interested in more than classification, however. They are also interested in seeing whether there are any predictable correlations between particular morphological characteristics or between morphological characteristics and other (syntactic or phonological) aspects of grammar. In other words, they seek to find whether there are any patterns to the kinds of morphology one finds in a language, and if there are, why those patterns might exist. For example, as *Whaley (1997: 131)* points out, isolating languages usually have rigidly fixed word orders. This correlation makes perfect sense: if a language has no morphological way of marking the *function* of noun phrases in a sentence, those functions must be signaled by the *position* of a noun phrase in a sentence.

Linguists such as *Joseph Greenberg (1963)* and *Joan Bybee (1985)* have looked at many languages and observed that there are several other correlations to be found. For example, Greenberg noted the two patterns in (51):

- (51) a. If a language has inflection, it will also have derivation.
 b. If a language has separate morphemes for number and case, and if both are either prefixes or suffixes, the number morpheme almost always occurs closer to the base than the case morpheme.

Observations such as these are called **implicational universals**. Based on observations of lots of languages, they are not true in every language, but they are true in a statistically significant number of languages.

Bybee (1985) also observed a statistically significant trend in the ordering of inflectional affixes in the languages of the world. What she noticed is that in languages with a number of different inflectional affixes on verbs, those affixes tended to come in a particular order. For example, if a language exhibits both tense and person/number affixes, the tense affix usually comes closer to the verb stem than the person/number affix. And if there are aspectual affixes, these tend to be closer to the verb stem than tense affixes.

We must still ask, however, why these particular correlations should exist. Bybee, for example, claims that the order of inflectional morphemes on verbs has something to do with their relevance to the verb itself. We might think of this in terms of the concepts of inherent and contextual inflection that we discussed in Section 6.6. For example, tense and aspect are inherent categories for verbs, but person and number are contextual: they signal agreement with one or more of a verb's arguments (its subject or object, for example). So inherent inflection comes closer to the verb stem than contextual inflection. The second of Greenberg's implicational universals might be explained in the same way. Number is an inherent inflection on nouns, but case is contextual, and inherent marking comes closer to the noun stem than contextual marking. Whether this is generally the case, however, is something that will require looking at many more languages.

7.5 Genetic and Areal Tendencies

In addition to classifying languages on the basis of specific structural characteristics that they display in their morphologies, we can look at typological patterns in a more global way. There are two ways to do this.

We can look at whether there are sorts of morphology that tend to be prevalent in particular language families or sub-families. And we may look at whether there are specific sorts of morphology that tend to be found in certain geographic areas even among languages that belong to different language families.

We can give several examples of genetic tendencies. If we look, for example, at compounding in two different branches of the Indo-European family, Italic (Romance) and Germanic, we can see an interesting pattern: although both branches make use of compounds, the sorts of compounds they favor are quite different. Germanic languages

like English tend to favor endocentric attributive and subordinate compounds like those in (52a), whereas Italic languages seem to prefer exocentric subordinate compounds, and have few attributive compounds (see Section 3.4 for these terms):

- (52) a. *English*
 endocentric attributive: dog bed, windmill
 endocentric subordinate: dishwasher, hand made
- b. *Italian*
 exocentric subordinate: lavapiatti ‘wash-dishes’ = ‘dishwasher’

Another example comes from the Bantu sub-family of the Niger-Congo languages. It is relatively rare in the languages of the world for inflectional morphology to be accomplished predominantly by prefixing. Nevertheless, as we mentioned above, there is a large concentration of such languages in the central and southern parts of Africa. Bantu languages frequently inflect nouns and verbs by adding prefixes. Similarly, although the vast majority of the languages of the world display some sort of inflection, many languages in the Sino-Tibetan language family are isolating. A nice illustration of this can be found in the WALS map that accompanies Chapter 26.

As for areal tendencies, Whaley (1997: 13) gives a fascinating example. He points out that three languages spoken in close proximity in the Balkan region of Europe all mark the definiteness of nouns by adding a suffix:

- (53) *Albanian* mik-u ‘friend-the’
 Bulgarian trup-at ‘body-the’
 Romanian om-ul ‘man-the’

What makes this example so interesting is that these three languages belong to completely different sub-families of Indo-European – Albanian forms its own branch, Bulgarian is Balto-Slavic, and Romanian is Italic – and none of the other languages in these three branches show definiteness with suffixes! Geographic proximity can be the only explanation for the distribution of this morphological trait.

Another example of a morphological pattern that is especially prevalent in a particular geographic region is verbal compounding. We saw in Chapter 3 that English rarely compounds two verbs (although there are a few examples like *stir-fry* or *slam-dunk*). In contrast, verbal compounds are not at all unusual in Asia, even in genetically unrelated languages. For example, although Japanese is a member of the Japonic family and Mandarin Chinese is a member of the Sino-Tibetan family, both display verbal compounds, as the examples in (54) show:

- (54) *Japanese* (Fukushima 2005: 570–85)
 naki-saken cry-scream ‘cry and scream’
- Mandarin* (Li and Thompson 1971: 55)
 mai-dao buy-arrive ‘manage to buy’

It would be interesting to explore the historical and social forces that lead languages in the same geographic area to develop similar morphological patterns, but we will not do so here (see [Heine 2014](#), though, for a brief treatment).

7.6 Typological Change

Although this is not a text about historical linguistics, in the context of typology it is important to point out that as languages change over time, their typological characteristics can change as well. I'll use English as an example here, as a fair amount is known about the evolution of the Indo-European language family in general and about the development and change of English morphology in particular. In [Chapter 6](#) we looked at changes to the inflectional system of English, specifically about the loss of inflection. To reprise, Old English had a far more complex inflectional system than modern English does, with grammatical gender, four cases, agreement between determiners, adjectives, and nouns, and a far more extensive system of verb inflection than we have now. By Middle English, much of this system had been lost. In typological terms, English moved farther away from being fusional and synthetic and closer to being isolating or analytic.

[Haselow \(2011: 216\)](#) shows the same trend over a much longer time span in looking at the development of Indo-European to the Germanic branch and ultimately to English. Historical linguists have hypothesized that Indo-European was highly fusional. Indo-European morphology frequently involved productive processes of internal vowel change (ablaut) in roots for purposes of both inflection and derivation. To these roots stem formatives and then inflections could be added. Some of these vowel changes survived into Germanic and ultimately into Old English, but over the course of time these ablaut processes had become unproductive: the only remnants we have are the strong verbs in Old English from which modern English's few 'irregular' verbs evolved. In terms of derivational morphology, a few verb and noun pairs like *sing* and *song* or *ride* and *road* survive, but it's not clear that speakers of modern English even see these as morphologically related. As the Indo-European system of vowel changes evolved and became unproductive, affixation, conversion, and compounding became more prominent derivational processes. Modern English, then, is typologically quite far from its Indo-European ancestor.

Summary

In this chapter we have surveyed five languages to see what morphological resources they make use of. Turkish is an agglutinative language that largely relies on suffixing. Mandarin Chinese is an inflectionally isolating language that makes heavy use of compounding to form new lexemes. Samoan makes use of a wide

range of different word formation processes to derive new lexemes, but is rather poor in inflection. Latin has heavily fusional inflection, and primarily derives new lexemes using prefixes and suffixes. And Nishnaabemwin is a polysynthetic language. We also looked at the traditional fourfold morphological classification of languages into isolating, agglutinative, fusional, and polysynthetic types, and at more useful typological tools such as the indexes of synthesis, of fusion, and of exponence, and at the distinction between head-and dependent-marking. Finally, we looked briefly at genetic and areal tendencies in morphological patterning.

Exercises

1. On the basis of the data below, try to classify these languages as isolating, agglutinative, fusional, or polysynthetic.

- a. *Swahili (Niger-Congo)* (Vitale 1981: 18)

Juma	a-li-wa-piga	watoto
Juma	3.SG.SUBJ-PST-3.PL.OBJ-hit	children
'Juma hit the children'		
Watoto	wa-li-m-piga	Juma
Children	3.PL.SUBJ-PST-3.SG.OBJ-hit	Juma
'The children hit Juma'		

- b. *Yay (Kra-Dai)* (Whaley 1997: 127)

mi ⁴	ran ¹	tua ⁴	ŋwa ¹	lew ⁶
not	see	CLASS	snake	CMPLT
'He did not see the snake' ⁷				

- c. *Musqueam (Salish)* (Suttles 2004: 28)

x^w-q^wé-nēc-t-əs
 inward-penetrate-bottom-TR-3_{TR}⁸
 '[She] punches holes in the bottom of it'

- d. *Old English* (Baugh and Cable 2013: 58)

On þyss-um	ēaland-e	cōm ūp sē
on this-NEUT.DAT.SG	island-NEUT.DAT.SG	came up the.MASC.NOM.SG
God-es	þēow	Augustinus
God-MASC.GEN.SG.	servant.MASC.NOM.SG	Augustine
and his gefēr-an.		
and his companion-MASC.NOM.PL		
'God's servant Augustine and his companions came up on this island.'		

7. The superscript numerals on the original Yay sentence in (b) above are tone markings.

8. _{TR} means 'transitive suffix'; 3_{TR} means '3rd person transitive subject'.

- e. *Náhuatl Puebla Sierra (Uto-Aztecan)* (Nida 1946: 171 – called Zacapoaxtla Aztec there)
 nan-čoka-to-skih-h
 2PL-CRY-DUR-COND-PL
 ‘you all would keep crying’

2. Consider the following paradigms from the Mayan language Tzutujil (Dayley 1985: 87):

waraam	‘to sleep’	
<i>Perfect</i>	1sg.	in warnaq
	2sg.	at warnaq
	3sg.	warnaq
	1pl.	oq warnaq
	2pl.	ix warnaq
	3pl.	ee warnaq
<i>Completive</i>	1sg.	xinwari
	2sg.	xatwari
	3sg.	(x)wari
	1pl.	xoqwari
	2pl.	xixwari
	3pl.	xeewari
<i>Incompletive</i>	1sg.	ninwari
	2sg.	natwari
	3sg.	nwari
	1pl.	noqwari
	2pl.	nixwari
	3pl.	neewari

Identify all the morphemes in the paradigms above. On the basis of your analysis, how would you classify Tzutujil using the traditional fourfold classification system?

3. On the basis of the data below, try to classify these languages as head-marking, dependent-marking, or double marking.
- Chechen (Nakh-Dagestanian)* (Nichols 1986: 60)
 de:-n a:xč a
 father-GEN money
 ‘father’s money’
 - Huallaga Quechua (Quechuan)* (Nichols 1986: 72)
 hwan-pa wasi-n
 John-GEN house-3
 ‘John’s house’
 - Abkhaz (Abkhaz-Adyghe)* (Nichols 1986: 60, from Hewitt 1979: 116)
 à-č ‘k’ən yə-y’ənə’
 the-boy his-house
 ‘the boy’s house’
4. Consider the forms below from the Yuman language Hualapai (data from Watahomigie, Bender, and Yamamoto 1982: 234). Divide the words into morphemes, gloss them, and then discuss whether possession in Hualapai is expressed by head-marking, dependent-marking, or double marking. If

you have trouble making a decision, discuss what further information about Hualapai would help you to decide.

nya hú'va	'my head'
ma mhú'ny	'your head'
nyihá hu'h	'his/her head'
nya qwáwva	'my hair'
ma mqwáwny	'your hair'
nyihá qwáwh	'his/her hair'

5. Review the sections of [Chapter 5](#) where we discussed morphological processes like reduplication, infixation, internal stem change, and templatic morphology. Do they present any problems for the traditional fourfold classification? Choose two examples from [Chapter 5](#) and discuss whether they are easily classified or not.
6. Look at the World Atlas of Language Structures Online (<https://wals.info>). At the WALS website click on "Features," and then click on article 24 "Locus of Marking in Possessive Noun Phrases." Which is more prevalent in the languages of the world, head-marking or dependent-marking? Now click on the accompanying map. Can you notice any areal tendencies in head-marking and dependent-marking in possessive noun phrases?
7. Look at the Universals Archive website (<http://typo.uni-konstanz.de/archive/intro/index.php>). Click on "Search" and at the search screen, type "morphology" in the box next to "Original," and then click on "Submit Query." You should get about 25 hits. From these hits, find five implicational universals.
8. Consider the following examples from the Otomanguean language Mixtec ([Macauley 1996: 79](#)):
 - a. a-ni-ka-žesámá-rí
 TEMP-CP-PL-eat-1
 'We already ate'
 - b. a-ni-ka-káǎ -ró xǐ maestro
 TEMP-CP-PL-talk-2 with teacher
 'You (PL) already talked with the teacher'

TEMP stands for temporal, CP for completive, and PL for plural. 1 and 2 stand for first person and second person respectively.

What problem does this example raise for Bybee's implicational universal?

8 Words and Sentences: The Interface between Morphology and Syntax

CHAPTER OUTLINE

KEY TERMS

valency
argument
active
passive
anti-passive
causative
applicative
noun
incorporation
clitic
phrasal verb
phrasal
compound

In this chapter you will learn about the intersection between morphology and syntax, which is the study of sentence structure.

- ◆ We will examine morphological processes like passivization and causativization that change the number of arguments of a verb.
- ◆ We will look at phenomena that sit on the borderline between word formation and syntax: clitics, phrasal verbs like *call up* in which the two parts can sometimes be separated in sentences, and compounds that contain whole phrases or even whole sentences.
- ◆ We will look at the competition between syntactic and morphological expression of grammatical concepts.

8.1 Introduction

As you've learned so far in this book, morphology is concerned with the ways in which words are formed in the languages of the world. Syntax, in contrast, is concerned with identifying the rules that allow us to combine words into phrases and phrases into sentences. Morphology and syntax, then, are generally concerned with different levels of linguistic organization. Morphologists look at processes of lexeme formation and inflection such as affixation, compounding, reduplication, and the like. Syntacticians are concerned, among other things, with phrase structure and movement rules, and rules concerning the interpretation of anaphors and pronouns. Nevertheless, there are many ways in which morphology and syntax interact, and indeed where the line between syntax and morphology is blurred.¹

We saw in [Chapter 6](#) that inflectional morphology is defined as morphology that carries grammatical meaning; as such it is relevant to syntactic processes. Case-marking, for example, serves to identify the syntactic function of an NP in a sentence. Inflectional markers like tense and aspect affixes identify clauses of certain types, for example, finite or infinitive, conditional or subjunctive. Person and number markers often figure in agreement between adjectives and the nouns they modify, or between verbs and their subjects or objects. In some sense, inflection can be viewed as part of the glue that holds sentences together.

In this chapter we will first look in more detail at several types of verbal morphology that affect sentence structure by changing what is called the valency of verbs. **Valency** concerns the number of **arguments** in a sentence, where arguments are noun phrases like the subject and object selected by the verb of the sentence. In [Section 8.3](#) we will look at the borderline between morphology and syntax. While it is usually clear to linguists which phenomena belong to which level of organization, the boundary between the two levels is not always crystal clear. There are cases in which derivational morphemes appear to attach to whole phrases, for example, or elements that seem not quite bound enough to be affixes, but not quite free enough to be viewed as independent words. Finally, in [Section 8.4](#) we will consider the formation of comparative and superlative adjectives in English as a case where both morphological and syntactic means of expression exist and where a complex set of conditions influences which type of expression is favored in any given environment.

8.2 Argument Structure and Morphology

Above, we defined the valency of a verb as the number of arguments it takes. **Arguments**, in turn, are defined as those phrases that are semantically necessary for a verb or are implied by the meaning of the verb.

1. As we will see in [Chapter 11](#), not all morphologists believe that there is a strict – or even fuzzy – separation between morphology and syntax.

Generally, arguments occur obligatorily with a verb, as the examples in (1) show:

- (1) a. Fenster snores.
 *Snores.
 b. Fenster devoured the pizza.
 *Fenster devoured.
 *Devoured the pizza.
 c. Fenster put the wombat in the bathtub.
 *Fenster put the wombat.
 *Fenster put in the bathtub.
 *Fenster put.

The verb *snores* has only one argument, its subject noun phrase. Verbs that have only one argument are traditionally called **intransitive**. The verb *devour* requires two arguments, its subject and object noun phrases. Two-argument verbs are **transitive**. And the verb *put* requires a subject, an object, and another phrase that expresses location. If a verb requires three arguments, it is traditionally called **ditransitive**.

The arguments of a verb are often, but not always, obligatory. For example, the verb *eat* must have a subject, and it can have an object, but the object is not necessary.

- (2) a. My goat eats tin cans.
 b. My goat eats.

Although it is optional we still consider the object *tin cans* to be an argument of the verb because the verb *eat* implies something eaten, even if that something is not overtly stated; notice that in (2b), we assume that the goat eats something (more specifically, we assume that the goat eats something foodlike, if we don't explicitly say otherwise, as we do in (2a)).

Of course, in English it is always possible to add prepositional phrases to a sentence that express the time or location of an action, the instrument with which it is done, or the manner in which it is done (*Fenster put the wombat in the bathtub with great care on Thursday ...*). These extra phrases are not necessary to the meaning of the verb, however, and are not arguments. Instead, they are called **adjuncts**. We will not need to talk about adjuncts here.

Valency-changing morphology alters the number of arguments that occur with a verb, either adding or subtracting an argument, making an intransitive verb transitive or a transitive verb intransitive, for example. English has some morphology that changes argument structure, but other languages, as we will see, have far more morphology of this sort.

8.2.1 Passive and Anti-Passive

The most obvious example of valency-changing morphology in English is the passive voice. Example (3) shows a pair of active and passive sentences in English:

- (3) a. Fenster bathed the wombat.
b. The wombat was bathed (by Fenster).

In the active sentence, the verb has two arguments, its agent (the one who does the action) and the patient or theme (what gets affected or moved by the action); the agent functions as the subject of the action, and the patient as the object. In the passive sentence, an agent is unnecessary. If it occurs, it appears in a prepositional phrase with the preposition *by*. The patient is the subject of the passive sentence. In effect there is no longer any object, and the passive form of the verb therefore has one fewer argument than the active form.

Part of what signals the passive voice in English is passive morphology on the verb. English passives are formed with the auxiliary verb *be* and a past participle, which is signaled for regular verbs by adding *-ed* to the verb base. Irregular verbs can form the past participle in a number of ways: by adding *-en* (*write ~ written*), by internal vowel change (*sing ~ sung*), or by internal vowel change and addition of *-t* (*keep ~ kept*).

Other languages also have morphological means to signal the change in argument structure in passive sentences. Example (4) shows an active and a passive sentence from West Greenlandic, and (5) an active–passive pair from Maori:

- (4) *West Greenlandic (Eskimo-Aleut)* (Fortescue 1984: 265)

- a. inuit nanuq taku-aat
people.REL² polar.bear see-3PL.3SG.INDIC
'The people saw the polar bear'
- b. nanuq (inun-nit) taku-niqar-puq
polar.bear (people-ABL) see-PASSIVE-3SG.INDIC
'The polar bear was seen (by the people)'

- (5) *Maori (Austronesian)* (Bauer 1993: 396)

- a. E koohete ana a Huia i a Pani
T/A scold T/A pers Huia DO pers Pani³
'Huia is scolding Pani'
- b. I koohete-tia a Pani e Huia
T/A scold-pass pers Pani by Huia
'Pani was scolded by Huia'

In (4b), you can see that the morpheme *niqar* makes a verb passive in West Greenlandic, and (5b) shows that a verb in Maori can be made passive by adding the suffix *-tia*.⁴ As was the case in English, the addition of these morphemes goes along with passive syntax, that is, making the

2. The relative case is the case of the transitive subject in West Greenlandic.

3. T/A means 'tense/aspect'; DO means 'direct object'.

4. Bauer (1993: 396–7) actually shows that there are several suffixes that make verbs passive in Maori.

patient/theme into the subject of the sentence, and making the agent optional. If the agent appears, it is marked with the ablative case in West Greenlandic, and by the preposition *e* in Maori.

Passive sentences are relatively familiar to speakers of English, but English has nothing like what is called the **anti-passive**. Like the passive, the anti-passive takes a transitive verb and makes it intransitive by reducing the number of its arguments. What's different, though, is which argument gets eliminated. For the passive, it's the transitive subject that disappears (or is relegated to a prepositional phrase or a case form other than that typical for subject), whereas for the anti-passive, it's the transitive object that disappears, as the example in (6) from Yidiñ shows:

- (6) *Yidiñ (Pama-Nyungan)* (Dixon 1977: 279)
- | | | |
|-----------------|------------------|-----------------|
| yĩṇu | buṇa | buga:-dĩṇ |
| this.ABSOLUTIVE | woman.ABSOLUTIVE | eat-ANTIPASSIVE |
- 'This woman is eating'

In Yidiñ, the anti-passive is marked on the verb by adding the suffix *:-dĩṇ*. Since Yidiñ is an ergative case-marking language (see Section 6.2), the subject of a transitive verb is in the ergative case. The subject of an intransitive verb is in the absolutive case, as you see in example (6). So while 'eat' is normally transitive in Yidiñ, you can see that it has become intransitive here.

As with the passive, it is also possible to express the 'missing' argument overtly, but in a case form other than that usually used for the direct object. Whereas in an active sentence, the direct object of a transitive verb is marked with the absolutive case, in an anti-passive sentence, the subject is absolutive and the object, if it appears, is either in the dative or the locative case:

- (7) *Dixon (1977: 277)*
- | | | |
|----------------|-----------------|------------|
| wagu:da | wawa:-dĩṇu | gudaga-nda |
| man.ABSOLUTIVE | saw-ANTIPASSIVE | dog-DATIVE |
- 'The man saw the dog'

While the translation of (7) makes it look like this sentence means exactly the same thing as an active sentence, this is only because there is no real way of capturing the nuance of this sentence in English, a language that lacks the anti-passive.

8.2.2 Causative and Applicative

Passive and anti-passive morphology signal a reduction in the number of arguments that a verb has. There are other sorts of morphology that signal that arguments have been *added* to a verb.

Causatives signal the addition of a new subject argument, which semantically is the causer of the action. If the verb has only one argument to begin with, the causative sentence has two, and if it has two to

begin with, the causative sentence has three arguments. Compare the Swahili sentences in (8) and (9):

- (8) *Swahili (Niger-Congo) (Vitale 1981: 158)*
- a. maji ya-me-chemka
water it-PER-boil
'The water boiled'
 - b. Badru a-li-chem-sh-a maji
Badru he-PST-boil-CAUSE water
'Badru boiled the water' (lit. 'caused the water to boil')
- (9) *Swahili (Vitale 1981: 156)*
- a. Halima a-li-ki-pika chakula
Halima she-PST-it-cook food
'Halima cooked the food'
 - b. Juma a-li-m-pik-ish-a Halima chakula
Juma he-PST-her-cook-CAUSE Halima food
'Juma caused Halima to cook food'

In (8a), the verb 'boil' has only one argument, its patient/theme. In (8b), along with the causative morpheme *-(i)sh*, an external causer argument is added as the subject of the sentence. Similarly, in (9a), the verb 'cook' has two arguments, an agent (*Halima*) and a patient ('food'); the agent is the subject of the sentence. When the causative suffix *-(i)sh* is added, a third argument (*Juma*) is added and it becomes the subject.

Applicative morphology, like causative morphology, signals the addition of an argument to the valency of a verb. But the added argument is an object, rather than a subject. We can again use Swahili for our example:

- (10) *Swahili (Vitale 1981: 44; Baker 1988: 393)*
- a. ni-li-pika chakula
I-PST-cook food
'I cooked some food'
 - b. ni-li-m-pik-i-a chakula Juma
I-PST-for him-cook-APPL food Juma
'I cooked some food for Juma'

The suffix *-i* signals that a second object (*Juma*) has been added to the verb.

Challenge

Aside from the passive, English is usually said to have little in the way of valency-changing morphology. Consider the following data, though:

- i. Fenster ate pickles.
- ii. Fenster over-ate.
- iii. *Fenster over-ate pickles.

What is the effect of the prefix *over-* on the valency of the verb *eat*?
Now consider sentences (iv–vi):

- iv. The plane flew.
- v. *The plane flew the field.
- vi. The plane over-flew the field.

Think about these examples, and try to think of other verbs formed with the prefix *over-* in English. Or better yet, try searching for them in COCA or the BNC. Does *over-* work like any of the valency-changing morphemes we have looked at in this chapter?

8.2.3 Noun Incorporation

There is one more way in which morphology interacts with the argument structure of verbs. Consider the data in (11) from Mapudungun:

- (11) *Mapudungun (Araucanian)* (Baker and Fasola 2009: 595)
- a. Ñi chao kintu-le-y ta chi pu waka
 my father seek-PROG-IND.3SG.SBJ the COLL cow
 ‘My father is looking for the cows’
 - b. Ñi chao kintu-waka-le-y
 my father seek-cow-PROG-IND.3SG.SBJ
 ‘My father is looking for the cows’

Sentences (11a) and (11b) mean precisely the same thing in Mapudungun. In (11a), the direct object ‘the cows’ is an independent noun phrase in the sentence, but in (11b), it forms a single compound-like word with the verb root ‘seek’. This sort of structure – where the object or another argument of the verb forms a single complex word with the verb – is called **noun incorporation**. Noun incorporation tends to occur in languages with polysynthetic morphology (see Section 7.4). In Mapudungun, the object noun follows the verb root in both the incorporated and the unincorporated forms. But this need not be the case, as the example in (12) from Mohawk shows:

- (12) *Mohawk (Iroquoian)* (Baker 1988: 20)
- a. Ka-rakv ne sawatis hrao-nuhs-a?
 3N-be.white DET John 3M-house-SUF
 ‘John’s house is white’
 - b. Hrao-nuhs-rakv ne sawatis
 3M-house-be.white DET John
 ‘John’s house is white’

As (12a) shows, the direct object follows the verb when it occurs independently in Mohawk, but it precedes the verb when it is incorporated.

There is much discussion among morphologists and syntacticians whether noun incorporation should be explained as a result of morphological rules or syntactic rules. We will return to this question in Chapter 10.

8.3 On the Borders

As we saw in the last section, one point of tangency between morphology and syntax occurs where morphology has an effect on the argument structure of verbs. There, it was clear that affixes – clearly morphological elements – can reduce or increase the number of arguments that a verb takes – traditionally a matter of syntax. What we will look at in this section, however, are cases where it is not so clear what belongs to morphology and what belongs to syntax – cases, in other words, that inhabit a sort of borderland between the two levels of organization.

8.3.1 Clitics

One of these borderland creatures is something that linguists call a **clitic**. Clitics are small grammatical elements that cannot occur independently and therefore cannot really be called free morphemes. But they are not exactly like affixes either. In terms of their phonology, they do not bear stress, and they form a single phonological word with a neighboring word, which we will call the host of the clitic. However, they are not as closely bound to their host as inflectional affixes are; frequently they are not very selective about the category of their hosts. Those clitics that come before their hosts are called **proclitics**, those that come after their hosts **enclitics**.

Two types of clitics are often distinguished: **simple clitics** and **special clitics**. Anderson (2005: 10) defines simple clitics as “unaccented variants of free morphemes, which may be phonologically reduced and subordinated to a neighboring word. In terms of their syntax, though, they appear in the same position as one that can be occupied by the corresponding free word.” In English, forms like *-ll* or *-d*, as in the sentences in (13), are simple clitics:

- (13) a. I’ll take the pastrami, please.
b. I’d like the pastrami, please.

In these sentences, *-ll* and *-d* are contracted forms of the auxiliaries *will* and *would*, and they occur just where the independent words would occur – following the subject *I* and before the main verb. Like affixes, they are pronounced as part of the preceding word. Unlike affixes, they do not select a specific category of base and change its category or add grammatical information to it. Contracted forms like *-ll* or *-d* in English will attach to any sort of word that precedes them, regardless of category:

- (14) a. The kid over there’ll take a pastrami sandwich.
b. No one I know’d want a pastrami sandwich.

In (14a) *-ll* is cliticized to the adverb *there*, and in (14b) *-d* is cliticized to the verb *know*.

Special clitics are phonologically dependent on a host, as simple clitics are, but they are not reduced forms of independent words. Compare the example in (15) from French:

- (15) a. Je vois Pierre.
I see Pierre.
- b. Je le vois.
I him see.
- c. *Je vois le.
I see him.

Although the object pronoun *le* in French is written as a separate word, it is phonologically dependent on the verb to its right; in other words, the object pronoun and the verb are pronounced together as a single phonological word. There is no independent word that means ‘him’ in French. So *le* and the other object pronoun forms in French are special clitics.

Clitics are of interest both to syntacticians and to morphologists precisely because they have characteristics both of bound morphemes and of syntactic units. Like bound morphemes, they cannot stand on their own. But unlike morphemes, they are typically unselective of their hosts and have their own independent functions in syntactic phrases.

8.3.2 Phrasal Verbs and Verbs with Separable Prefixes

Also inhabiting the borderland between morphology and syntax are phrasal verbs in English and verbs with separable prefixes in German and Dutch. **Phrasal verbs** are verbs like those in (16) that consist of a verb and a preposition or particle:

- (16) call up ‘telephone’
chew out ‘scold’
put down ‘insult’
run up ‘accumulate’

Frequently, phrasal verbs have idiomatic meanings, as the glosses in (16) show, and in that sense they are like words. In terms of structure, the combination of verb and particle/preposition might seem like another sort of compound in English. Remember, however, that one of the criteria for distinguishing a compound from a phrase in English (see Section 3.4) was that the two elements making up compounds could not be separated from one another. We cannot take a compound like *dog bed* and insert a word to modify *bed* (e.g., **dog comfortable bed*). In contrast, however, the two parts of the phrasal verb can be, and sometimes must be, separated:

- (17) a. I called up a friend.
b. I called a friend up.
c. I called her up.
d. *I called up her.

When the object of the verb is a full noun phrase, the particle can precede or follow it. In the former case it is adjacent to its verb, but in the latter case it is separated from the verb. And when the object is a pronoun, the particle must be separated from the verb. So do we consider phrasal verbs to be a matter of study for morphologists, or do we leave them to syntacticians? There is no set answer to this question.

A similar issue arises with what are called **separable prefix verbs** in Dutch. Consider the examples in (18):

(18) Booij (2002: 205)

- a. ... dat Hans zijn moeder opbelde ~ Hans belde zijn moeder op
that Hans his mother up.called ~ Hans called his mother up
- b. ... dat Jan het huis schoonmaakte ~ Jan maakte het huis schoon
that Jan the house clean.made ~ Jan made the house clean
- c. ... dat Rebecca pianospeelde ~ Rebecca speelde piano
that Rebecca piano.played ~ Rebecca played piano

Like phrasal verbs in English, separable prefix verbs in Dutch often have idiomatic meanings. For example, *opbellen*, like English 'call up', means 'to telephone'. Each separable prefix verb in Dutch consists of a verb preceded by a word of another category; in (18), *op* is a preposition, *schoon* is an adjective, and *piano* is a noun. These words therefore look a bit like prefixed words or perhaps compounds. But there's a difference.

To understand the examples in (18) you need to know that Dutch exhibits different word orders in main clauses than in subordinate clauses. In main clauses the main verb is always the second constituent in the clause. If the subject is first, the main verb comes right after it. But in subordinate clauses, the main verb always comes last. The examples in (18) show that when the verbs *opbellen* 'call up', *schoonmaken* 'make clean', and *pianospeelen* 'play piano' occur in a subordinate clause, those complex verbs come at the end of the clause. The first elements *op*, *schoon*, and *piano* occur attached to the verb, almost like prefixes. However, when the verb is used in a main clause, the verb itself occurs after the subject, but its first element appears separated from it at the end of the sentence. So unlike normal prefixes, these elements sometimes are not attached to the verbs with which they normally form a unit. Are separable prefix verbs a matter for morphologists or for syntacticians? Again, there is no easy answer to this question, as they lie on the border between the two.

8.3.3 Phrasal Compounds

Our final example of a phenomenon that is neither clearly syntactic nor clearly morphological is called a **phrasal compound**. A phrasal compound is a word that is made up of a phrase as its first element, and a noun as its second element. Phrasal compounds can be found in many of the Germanic languages, including English, Dutch, and German:

- (19) a. *English* (Harley 2009)
stuff-blowing-up effects

bikini-girls-in-trouble
genre comic-book-and-science-fiction fans

- b. *Dutch* (Hoeksema 1988)
luch of ik schiet humor
'laugh or I shoot humor'
- c. *German* (Toman 1983: 47)
die Wer war das Frage
'the who was that question'

On the one hand, phrasal compounds pass one of the acid tests for compounding: it is impossible to insert a modifying word in between the phrase and the head of the compound:

- (20) a. *stuff-blowing-up exciting effects
b. exciting stuff-blowing-up effects

On the other hand, the first elements are clearly phrases, or even whole sentences, as the example in (21) shows:

- (21) God-is-dead theology

And phrases and sentences are the subject matter of syntax. Again, it is no easy question to decide whether phrasal compounds are the subject of morphology or of syntax. Indeed, it would be reasonable to conclude that they should be of interest to both morphologists and syntacticians.

Challenge

Consider the forms below:

wife-and-motherdom
do-it-yourself-er
old-fartery
one-step-behindhood
dog-in-the-mangerish
get-even-with-themism
I-can-do-it-too-ness
anti-business-as-usual
ex-friends-with-benefits
post-fall-of-Baghdad
pre-shock-and-awe

All of these are words that were found in COCA (taken from Bauer, Lieber, and Plag 2013: 513–14). Discuss the internal structure of these words and then consider what sorts of issues they raise for the interaction of morphology and syntax.

8.4 Morphological *versus* Syntactic Expression

The final topic we will consider in this chapter is that of morphological versus syntactic expression of particular grammatical distinctions. I'll illustrate what I mean with a comparison of the future tense in English and French. In English, the future is signaled syntactically, using a combination of the modal auxiliary *will* plus the uninflected form of the verb, as we see in (22a). In French, however, the future tense is expressed using an inflectional ending, as in (22b):

- (22) a. *English*
 I will speak.
 He will arrive.
- b. *French*
 Je parl-er-ai.
 I speak-FUT-1.SG
 Il arriv-er-a.
 he arrive-FUT-3.SG

One way of describing the difference between English and French in this regard is that the future is expressed periphrastically in English but morphologically in French. We can refer to a way of expressing a concept as **periphrastic** when we use a syntactic construction consisting of two or more separate words to express that concept.

Another example can be illustrated with causatives. We saw earlier that in Swahili verbs are inflected with a specific morpheme to mean 'cause to verb':

- (23) Badru a-li-chem-sh-a maji
 Badru he-PST-boil-CAUSE water
 'Badru boiled the water' (lit. 'caused the water to boil')

In English we do not have a morphological causative, but we can express causativity periphrastically:

- (24) I made the water boil.

Note that for this particular verb, we can also use the verb alone with the causative meaning, as in the sentence *I boiled the water*. But we have no specific morphological means of making a causative verb.

Interestingly, English displays at least one example where a particular grammatical concept can be expressed either morphologically or periphrastically, namely the inflection of adjectives for the comparative and superlative. When comparative and superlative are expressed morphologically, we use the suffixes *-er* and *-est*. When we express them periphrastically, we use *more* for the comparative and *most* for the superlative.

(25)	<i>adjective</i>	<i>comparative</i>	<i>superlative</i>
	red	redder	reddest
	pure	purier	purest
	active	more active	most active
	honest	more honest	most honest

Challenge

Consider the list of adjectives below and try to decide whether you prefer to form the comparative and superlative morphologically or periphrastically:

wide	intelligent	horrible
gentle	direct	defunct
serene	stunning	ill
bumpy	official	Spartan
beautiful	difficult	famous
fake	meager	bold

What factors seem to influence your choice between the morphological and the periphrastic comparative and superlative? Are there any adjectives for which both morphological and periphrastic forms seem possible? Are there any for which neither the morphological nor the periphrastic forms sounds OK?

Morphologists have found that there are a number of factors that influence whether speakers prefer the morphological or the periphrastic comparative and superlative. Phonological factors that influence the choice have to do with the number of syllables in the base adjective and its stress pattern. One-syllable adjectives favor the morphological realization and adjectives of three or more syllables prefer the periphrastic form (e.g., *red*, *redder*, *reddest* as opposed to *difficult*, *more difficult*, *most difficult*). Two-syllable adjectives can usually take the morphological comparative and superlative if their second syllable is unstressed, but if the second syllable is stressed, the periphrastic form is preferred (e.g., *happy*, *happier*, *happiest* versus *direct*, *more direct*, *most direct*). Morphological factors that influence the choice include whether or not the adjective is simple or complex. For complex adjectives, the periphrastic form is usually preferred (e.g., *active*, *more active*, *most active*). And syntactic factors include whether the adjective is followed by a preposition and a comparative clause or not in its syntactic context: consider, for example, whether you prefer *more proud* or *prouder* in a context like *I am _____ of Fenster than of Letitia*. What is most interesting about the comparative and superlative in English is that there are no hard and fast rules. Even if speakers tend to prefer one or the other comparative and superlative forms for particular adjectives, many adjectives can form their

comparatives and superlatives either way. So if you check COCA, you might find *more red* along with *redder*, for example.

Summary

In this chapter we have investigated the relationship between morphology and syntax. We have seen that there are ways in which morphology affects the syntax of sentences, by either reducing or increasing the number of arguments a verb may appear with. We have also looked at cases where it is not entirely clear whether a phenomenon is a matter of morphology or of syntax or of both. Among these phenomena, we find clitics, phrasal verbs, separable prefix verbs, and phrasal derived words and compounds. And we have looked briefly at an example where morphological and syntactic expression are both possible. What such phenomena really show us is that morphology and syntax are often intimately intertwined, and often morphologists must investigate both levels of organization to really understand how a language works. We will see in [Chapter 10](#) that phenomena like the ones we've looked at in this chapter raise serious questions for theorists, and have been the matter of much discussion.

Exercises

1. Consider the sentences below and discuss how the passive is formed in Swahili. Use sentences (8)–(10) in this chapter to help you gloss these sentences. [Vitale \(1981: 23, 31\)](#)
 - a. Juma a-li-fungua mlango
'Juma opened the door'
 - b. mlango u-li-fungu-liwa
'The door was opened'
2. Consider the prefix *out-* in English:
 - a. Fenster ran.
 - b. *Fenster ran Letitia.
 - c. Fenster outran Letitia.
 - d. *Fenster outran.

Describe the effect that the prefix *out-* has in sentences (a)–(d). Now, think of other verbs formed with *out-* in English. Does *out-* have a consistent effect on the argument structure of verbs it attaches to?
3. Although phrasal compounds may seem somewhat exotic to you, they appear not infrequently in journalistic writing, especially in headlines, and in more informal writing, for example, on the sports pages or in feature writing. Choose your favorite newspaper and try to find two examples of phrasal compounds. Share your examples with classmates and try to analyze what sorts of phrases can occur as the first elements of your compounds.

4. Discuss the difference in argument structure and in verbal morphology between the pairs of sentences below:

a. *Malagasy (Austronesian)* (Keenan and Polinsky 1999: 604)

- i. mijaly Rabe
suffers Rabe
'Rabe suffers'
- ii. mampijaly an-dRabe Rasoa
makes-suffer ACC-Rabe Rasoa
'Rasoa makes Rabe suffer'

b. *Chichewa (Niger-Congo)* (Mchombo 1999: 506)

- i. Kalúlú akuphíká maúngu
Hare is cooking pumpkins
'The hare is cooking pumpkins'
- ii. Kalúlú akuphíkíra mkángó maúngo
Hare is cooking lion pumpkins
'The hare is cooking pumpkins for the lion'

5. Consider the English sentences below:

- a. The water boiled.
- b. Fenster boiled the water.
- a. The tomatoes grew.
- b. Letitia grew the tomatoes.
- a. The door opened.
- b. Roddy opened the door.

First discuss the difference in argument structure between the (a) sentences and the (b) sentences. Then compare these sentences to the Swahili sentences in (8) (see p. 170). Discuss the differences between the Swahili sentences and the ones here.

6. Musqueam (Salish) exhibits an interesting set of morphologically related verbs that differ in valency. Analyze the following forms and describe both the affixes used and their effects on argument structure (Suttles 2004: 235) (note that the first two examples are a bit different from the second two):

háy	'finish'	shá·y	'finished'	shá·y stəx ^w	'have it finished'
lólqt	'dip it'	slélq	'in the water'	slélqstəx ^w	'keep it in the water'
ǰé·t ^h	'measure it'	sǰ eʔé·t ^h	'measured, marked'	sǰ eʔé·t ^h stəx ^w	'blaze (as a trail), designate (as a time)'
θáyt	'fix it'	sθəθəy	'right'	sθəθəystəx ^w	'keep it on course'

7. For each of the adjectives below, check in COCA whether you can find the morphological comparative and superlative, the periphrastic comparative and superlative, or both. For example, for the adjective *pure*, you'd look for *purier*, *more pure*, *purest*, and *most pure*. See what you find and discuss the factors that seem to influence the choice of the morphological and periphrastic forms.

weird
winning (as in *a winning team*)
happy
famous

9 Sounds and Shapes: The Interface between Morphology and Phonology

CHAPTER OUTLINE

KEY TERMS

allomorphy
assimilation
epenthesis
underlying representation
vowel harmony
syllable
onset
nucleus
coda
rhyme
palatalization
lexical strata

In this chapter we will learn about the intersection between morphology and phonology, which is the study of the sound structure of languages.

- ◆ We will learn that some morphemes exhibit allomorphs, that is, phonologically distinct variants.
- ◆ We will learn how to analyze both phonologically predictable and unpredictable allomorphy.
- ◆ We will learn about how syllable structure and morphology can interact.
- ◆ And we will consider the nature of lexical stratification, where different types of phonological behavior characterize different parts of the morphology of a language.

9.1 Introduction

Phonology is the area of linguistics that is concerned with sound regularities in languages: what sounds exist in a language, how those sounds combine with each other into syllables and words, and how the prosody (stress, accent, tone, and so on) of a language works. Phonology interacts with morphology in a number of ways: morphemes may have two or more different phonological forms whose appearance may be completely or at least partly predictable. Some phonological rules apply when two or more morphemes are joined together. In some languages morphemes display different phonological behavior depending on whether they are native to the language or borrowed into it from some other language. Some morphological rules may depend on the syllable structure of a language. In this chapter we will explore the various ways in which phonology interacts with morphology.

In this chapter we will frequently make use of phonetic transcriptions, so you may want to review the IPA before you begin reading it. We will also make use of terminology which classifies sounds by their place of articulation (labial, dental, alveolar, and so on) and by their manner of articulation (voiced vs. voiceless, stop, fricative, liquid, and so on). You can find summaries of this terminology in the charts at the beginning of the book.

9.2 Allomorphy and Morphophonological Rules

Allomorphs are phonologically distinct variants of the same morpheme. By phonologically distinct, we mean that they have similar but not identical sounds. And when we say that they are variants of the same morpheme, we mean that these slightly different-sounding sets of forms share the same meaning or function. For example, the negative prefix *in-* in English is often pronounced *in-* (as in *intolerable*), but it is also sometimes pronounced *im-* or *il-* (*impossible*, *illegal*), as English spelling shows. Since all of these forms still mean ‘negative’, and they all attach to adjectives in the same way, we say that they are allomorphs of the negative prefix. Another example you’ve already seen is the regular past tense in English. Although the regular past tense in English is always spelled *-ed*, it is sometimes pronounced [t] (*packed*), sometimes [d] (*bagged*), sometimes [əd] (*waited*).¹ Still all three phonological variants designate the past tense. Similarly, the plural morpheme in Turkish sometimes appears as *-lar* and sometimes as *-ler*, so Turkish has two allomorphs of the plural morpheme.

As we will see below, in many cases, it is phonologically predictable which allomorph appears where; sometimes, however, which

1. Or [ɹd] in some dialects.

allomorph appears with a particular base is unpredictable. For example, we will see that it is usually possible to predict the form of the regular allomorphs of the English past tense morpheme, but there are quite a few verbs whose past tenses are irregular (e.g., *sang*, *flew*, *bought*).

9.2.1 Predictable Allomorphy

Let's look more closely at the prefix *in-* in English. As the examples in (1a) show, it frequently has the form *in-*. However, sometimes it appears as *im-*, *il-*, or *ir-*, as the examples in (b) and (c) show. And if you think about sound rather than spelling, it can also be pronounced [ɪŋ-], as the examples in (1d) show:

- (1) a. inalienable
intolerable
indecent
b. impossible
c. illegal
irregular
d. incongruous [ɪŋkŋɡruəs]
incoherent [ɪŋkəhiənt]

The various allomorphs of the negative prefix *in-* in English are quite regular, in the sense that we can predict exactly where each variant will occur. Which allomorph occurs depends on the initial sound of the base word. For vowel-initial words, like *alienable*, the [ɪn-] variant appears. It appears as well on words that begin with the alveolar consonants [t, d, s, z, n]. On words that begin with a labial consonant like [p], we find [ɪm-]. Words that begin with [l] or [ɹ] are prefixed with the [ɪl-] or [ɪɹ-] allomorphs respectively, and words that begin with a velar consonant [k], are prefixed with the [ɪŋ-] variant. What you should notice is that this makes perfect sense phonetically: the nasal consonant of the prefix matches at least the place of articulation of the consonant beginning its base, and if that consonant is a liquid [l, ɹ] it matches that consonant exactly. This allomorphy is the result of a process called **assimilation**. Generally speaking, assimilation occurs when sounds come to be more like each other in terms of some aspect of their pronunciation.

If you have studied a bit of phonology, you know that regularities in the phonology of a language can be stated in terms of phonological rules. Phonologists assume that native speakers of a language have a single basic mental representation for each morpheme that we call the **underlying representation**. Regular allomorphs are derived from the underlying representation using phonological rules. For example, since the English negative prefix *in-* is pronounced [ɪn] both before alveolar-initial bases (*tolerable*, *decent*) and before vowel-initial bases (*alienable*), whereas the other allomorphs are only pronounced before specific consonant-initial bases, phonologists assume that our mental representation of *in-* is [ɪn] rather than [ɪɹ], [ɪl], or [ɪŋ]; often (but not always, as we will see below) the underlying form of a morpheme is the form that has

the widest surface distribution.² When the underlying form is prefixed to a base beginning with anything other than a vowel or alveolar consonant, the following phonological rule derives the correct allomorph:

- (2) *Place assimilation*: a nasal consonant assimilates to the place of articulation of a following consonant, and to the place and manner of articulation of the consonant if it is a liquid.

Phonologists use different forms of notation to express the above rule in a more succinct fashion, but we'll restrict ourselves to informal statements of rules here.

This sort of assimilation – called **place assimilation** – is not unusual in the languages of the world. We find something similar in Zoque (Mixe-Zoquean) (Nida 1946/1976: 21):

- (3) pama 'clothes' ?əs mpama 'my clothes'
 kayu 'horse' ?əs ɲkayu 'my horse'
 tuwi 'dog' ?əs ntuwi 'my dog'

As the examples in (3) show, the possessive prefix is a nasal consonant that has three different allomorphs. Which allomorph is prefixed depends on the place of articulation of the noun it attaches to. In Zoque, we might say that part of forming the possessive of a noun involves prefixing an underlying nasal consonant which undergoes a phonological rule that assimilates it to a following consonant.

Another example of a predictable form of allomorphy is the formation of the regular past tense in English. In Chapter 2, we looked at the past tense in English in the context of figuring out what the mental lexicon looks like. We can now go into its formation in somewhat more detail. Consider the data in (4), which shows two of the three allomorphs of the regular past tense:

- (4) a. *Verbs whose past tense is pronounced [t]*
 slap, laugh, unearth, kiss, wish, watch, walk
 b. *Verbs whose past tense is pronounced [d]*
 rub, weave, bathe, buzz, judge, snag, frame, can, bang, lasso, shimmy

The regular past tense in English illustrates a different sort of assimilation, called **voicing assimilation**, where sounds become voiced or voiceless depending on the voicing of neighboring sounds. The verbs that take the past tense allomorph [t] all end in voiceless consonants: [p, f, θ, s, ʃ, tʃ, k]. Those that take the [d] allomorph, all end either in a voiced consonant [b, v, ð, z, ʒ, dʒ, g, m, n, ŋ, etc.] or in a vowel (and vowels are typically voiced, of course). Why just this distribution? Clearly, the past tense morpheme has come to match the voicing of the final segment of the verb base: verbs whose last segment is voiceless take the voiceless variant.

2. If you study phonology further, you will find that this is somewhat of a simplification, but for our purposes, it is good enough.

There is one allomorph of the past tense we haven't covered yet. Consider what happens if the verb base ends in either [t] or [d]:

- (5) *Verbs whose past tense is pronounced [əd]*
defeat, bond

Here, a process of **epenthesis** is at work. Epenthesis is a process that inserts a segment into a word. Here, a schwa is inserted to separate [d] of the past tense from the alveolar stop at the end of the verb. Again, this makes perfect sense phonetically; if the [d] allomorph were used, it would be hard to distinguish from the final consonant of the verb root.

What is the underlying form of the past tense morpheme in English? As I indicated before, it is often a good strategy to assume that the allomorph with the widest distribution is the underlying form. But there is something else to consider as well. Phonologists typically assume that the underlying form of a morpheme must be something from which all of the other allomorphs can be derived using the simplest possible set of rules. In this case, the allomorph [d] has the widest distribution, because it occurs with all voiced consonants except [d], and with all vowel-final verb stems. And if we assume that the underlying form of the regular past tense is [d], we need only two simple rules to derive the other allomorphs:

- (6) *The Past Tense Rule*
- If the verb stem ends in [t] or [d] (the alveolar stops), insert [ə] before the past tense morpheme (e.g., defeated [dəfɪt + d] → [dəfɪt + əd]).
 - Assimilate [d] to the voicing of an immediately preceding consonant (e.g., licked [lɪk + d] → [lɪk + t]).

Challenge

Rather than take my word for it that choosing the allomorph [d] as the underlying representation of the past morpheme yields the simplest set of rules, construct an argument that the set of rules in (6) really is simpler than alternatives. To do so, first suppose that we had chosen [t] instead of [d] as the underlying morpheme. Try to state informally what the rule(s) would have to be to derive the other allomorphs of the past tense. Then suppose that we'd chosen [əd] as the underlying representation. What would the rules have looked like then? Now compare the rules to each other and discuss which set is simplest.

A third example of regular and predictable allomorphy comes from Turkish. As we've seen, in Turkish, virtually every morpheme, derivational and inflectional alike, has more than one allomorph. For example, the plural morpheme has the allomorphs *-ler* and *-lar*, and the genitive suffix has the allomorphs *-in*, *-un*, *-m*, and *-ün*.³ The reason for this is that

3. Orthographic <ü> is the high front rounded vowel /y/ in IPA.

Turkish displays a process of **vowel harmony** whereby all non-high vowels in a word have to agree in backness, and all high vowels in both backness and roundness. When suffixes are added to a base, they must agree in the relevant vowel characteristics with the preceding vowels of the base:

(7) *Turkish (Turkic)* (Lewis 1967: 29ff)

	‘hand’	‘measure’	‘evening’	‘fear’
Abs. pl.	el-ler	ölçü-ler	akşam-lar	korku-lar
Gen. sg.	el-in	ölçü-n-ün	akşam-ın	korku-n-un

Since the roots of the nouns *el* ‘hand’ and *ölçü* ‘measure’ have front vowels, the plural suffix must agree with them in frontness, so the *-ler* allomorph appears. On the other hand, *akşam* ‘evening’ and *korku* ‘fear’ have back vowels, and the *-lar* allomorph appears. Since the genitive ending has a high vowel, the vowel harmony is more complicated. If the noun root consists of vowels that are front and non-round, we find the genitive allomorph with a front, non-round vowel, that is, *-ın*. Similarly, if the root contains front, round vowels, so does the suffix; so *ölçü* gets the front round allomorph *-ün*. Roots with back non-round vowels like *akşam* take the *-m* allomorph, and roots with back round vowels like *korku* take the *-un* allomorph.

We might ask in this case what the underlying form of these affixes is, and here it’s a bit difficult to pick one of the existing allomorphs as our choice. For example, neither plural allomorph *-ler* nor *-lar* has a wider distribution than the other. One possibility that we might consider, then, is that the mental representation of the plural morpheme in Turkish is something like what we find in (8):

$$(8) \quad \text{Turkish plural morpheme} = -1 \begin{bmatrix} \text{V} \\ \text{non-high} \\ \text{non-round} \end{bmatrix} \text{r}$$

What the representation in (8) says is that the underlying form of the plural morpheme has an ‘underspecified’ vowel, that is, a vowel that is non-high and non-round, but inherently neither front nor back. Part of the rule of vowel harmony in Turkish might then say that a nonhigh, non-round vowel in a suffix comes to match the backness of the vowels in a root that precedes it.

Challenge

We have proposed an underlying form for the plural morpheme in Turkish. Now propose one for the genitive suffix and try to give an informal statement of the rule of vowel harmony that gives rise to the different allomorphs.

We can give one more example of predictable allomorphy from Turkish, which affects consonants rather than vowels. Let’s look at a bit

more data from Turkish (Myers and Crowhurst 2006 (www.laits.utexas.edu/phonology/turkish/index.html)):

(9)		<i>Nom.Sg</i>	<i>Dat.Sg</i>	<i>Nom.Pl</i>
	'stalk'	sap	sapa	saplar
	'container'	kap	kaba	kaplar

The examples in (9) are interesting because the forms look straightforward until we get to the Dative Singular of the word 'container'. If we look only at the Nominative forms of these words we might think that the underlying forms of the roots are *sap* and *kap* respectively, that is, that the underlying roots are identical to the nominative singular forms. However, when we look at the Dative form, we find that 'container' has a voiced consonant [b] preceding the dative suffix, where 'stalk' maintains its voiceless [p]. How can we explain why the two words behave differently in the Dative form? If we assume that the root of 'container' underlyingly ends in [b] rather than [p] (so the underlying roots are /*sap*/ and /*kab*/ respectively), we can say that voiced consonants become voiceless either word finally or before consonants, but remain voiced when they occur between vowels. Note that in this case, the underlying form of the word 'container' is not equivalent to any of the surface forms that we see in the data in (9).

So far, we have looked at relatively simple examples of allomorphy, but more complex allomorphy can and does occur in the languages of the world. Consider the data below from the Austronesian language Balantak (data from Busenitz and van den Berg 2012: 14–15). The forms illustrate the forms of the actor voice irrealis prefix:⁴

(10)	a.	mang-ala	'to take, get'
		meng-keke	'to dig'
		ming-ili	'to buy'
		mong-gopot	'to sort, arrange'
		mung-kukudi	'to peel'
	b.	mam-bala	'to fence in'
		mem-pepel	'to hammer'
		mim-pipir	'to splatter'
		mom-popok	'to cut'
		mum-buani	'to fish with a net'
	c.	man-taring	'to cook'
		men-deer	'to spread'
		min-sikoop	'to shovel'
		mon-sosop	'to suck'
		mun-tunu	'to burn'
	d.	manga-wawau	'to make'
		menge-leelo'	'to call, invite'
		mingi-nika'	'to marry'
		mongo-yoong	'to shake'
		mungu-rudus	'to pick'

4. Orthographic <ng> is phonetically [ŋ].

The examples in (10) show that there appear to be many allomorphs for the irrealis actor focus morpheme in Balantak, or to put it differently, that several phonological processes can affect that morpheme. First, it is apparent that the vowel in the prefix must be the same as the first vowel in the verb base. This is, of course, another case of vowel harmony. The examples in (10b) and (10c) show us that the final consonant of the prefix assimilates to the initial consonant of the base if it is an obstruent (i.e., a stop or a fricative). Finally, the examples in (10d) show us that an epenthetic vowel is inserted at the end of the prefix if the base begins with a sonorant and that this vowel harmonizes with the first vowel in the base. What, then, would the underlying form of all these allomorphs be? The data in (10a) suggest that the underlying form should be consonant-final rather than vowel-final, as the consonant-final allomorphs have the widest distribution. Further, the data suggest that the final consonant of the prefix should be *ng* [ŋ] because this consonant occurs both before velar-initial bases and vowel-initial bases (again, that means that it has the widest distribution). The consonant [m] occurs only before labials, and [n] only before alveolars. As for the vowel, there seems to be no argument for preferring one surface vowel over the other, so van den Berg and Busenitz simply show the underlying form with an unspecified vowel, symbolized by V. The underlying form, then, is *mVng-*.

9.2.2 Unpredictable or Partially Predictable Allomorphy

As we've seen above, some allomorphy is regular enough to be captured by phonological rules. But not all allomorphy is regular. Take, for example, past tenses of verbs in English. We have already looked at the regular past tense. Every native speaker or student of English knows that there are also quite a few verbs that don't form the past tense by adding *-ed*. Consider Table 9.1, which gives a selection of examples.

Table 9.1 (based on classes in Huddleston and Pullum 2002)

	Infinitive	Irregular past	Pattern
1	burn	burnt	devoicing of suffix
2	keep	kept	vowel shortening
3	hit	hit	no change
4	feel	felt	vowel shortening with devoicing of suffix
5	bleed	bled	vowel shortening and no suffix
6	leave	left	devoicing of stem consonant
7	sing	sang	vowel ablaut (i ~ æ)
8	win	won	vowel ablaut (i ~ ʌ)
9	fight	fought	vowel ablaut (ai ~ ɔ)
10	come	came	vowel ablaut (ʌ ~ e)

If you think back to [Chapter 2](#), when we discussed the mental lexicon, we suggested that irregular past tense allomorphs might simply be stored in the mental lexicon, and not derived by rules. So speakers of English would have a lexical entry for the verb root *sing*, and along with it an associated entry for past tense *sang*. It is possible, though, that things are a bit more complicated. Think back to the experiment in [Section 6.3](#) where you asked a number of friends to make the past tense of the hypothetical verb *gling*. Probably a significant number of them offered either *glang* or *ghung*. Since this is not a real verb, clearly they didn't have a past tense stored for it. Rather, they must have been making use of some sort of pattern to create these forms. In English there happen to be quite a few verbs whose present and past tenses show the same *i ~ æ* alternation as *sing* or the *i ~ ʌ* alternation of *win*. There appears to be an abstract pattern that speakers are tapping into here that relates a present tense with [i] to a past tense with [æ] or [ʌ] if the verb ends in a nasal or a nasal plus some other consonant (e.g., like *swim*, *ring*, *sting*, *win*, *stink*). Psycholinguists continue to work towards figuring out the exact nature of such patterns.

The example of unpredictable allomorphy we looked at above concerns English inflection. Let's look at another example that has to do with derivation. Consider the forms in (11):

- | | | | |
|-------------------|-----------------|--------------|--------------------------|
| (11) a. designate | ['dɛ.zɪg.nɜrt] | designation | [d.ɛzɪg.'neɪ.ʃ + ʌn] |
| b. unionize | ['ju.njə.naɪz] | unionization | [ju.njə.naɪ.'z + eɪ.ʃʌn] |
| c. prosecute | ['pɹɑ.sə.kjʊt] | prosecution | [pɹɑ.sə.'kju.ʃ + ʌn] |
| d. resolve | [ɹɔ.'zɒlv] | resolution | [ɹɛ.zə.'l + u.ʃʌn] |
| e. expedite | ['ɛk.spə.dart] | expedition | [ɛk.spə.'dɪ + ʃ.ʌn] |
| f. define | [də.'fɑɪn] | definition | [dɛ.fə.'n + ɪ.ʃʌn] |
| g. absorb | [əb.'zɔrb] | absorption | [əb.'zɔrp. + ʃʌn] |
| h. circumcise | ['səɪ.kʌm.saɪz] | circumcision | [səɪ.kəm.'sɪ.ʒ + ʌn] |
| i. decide | [də.'saɪd] | decision | [də.'sɪ.ʒ + ʌn] |

All of the verbs in the lefthand column have noun forms with the suffix *-tion*. But if you compare the transcriptions of the verbs and nouns carefully, you will see that both the verb bases and the derivational affix have various allomorphs. For example, the suffix seems to be *-ʌn* in (11a and c) but *-eɪʃʌn* in (11b). It looks like *-uʃʌn* in (11d), but *-ɪʃʌn* in (11f), and *-ʃʌn* in (11g). In (11a, c, and e) the [t] at the end of *designate*, *prosecute*, and *expedite* seem to have changed to [ʃ], the [v] at the end of *resolve* seems to have disappeared, and the [b] at the end of *absorb* has changed to [p]. And if you look carefully at many of these forms, the stress pattern on the derived noun is different from that of its verb base. In other words, there is quite a complicated pattern of allomorphy associated with this suffix.

Is it predictable? Parts of it are. For example, if a verb ends in [v] and has a derived noun with the *-tion* suffix, it will always lose its [v] and the suffix will be pronounced *-ution* (think about the derived nouns for *dissolve*, *absolve*, *revolve*, etc.). Similarly, if a verb ends in [t] and takes the *-ion* suffix, the [t] will become [ʃ]. And if a verb ends in [z] or [d] and takes the *-tion* suffix, those consonants will become [ʒ]. Since the

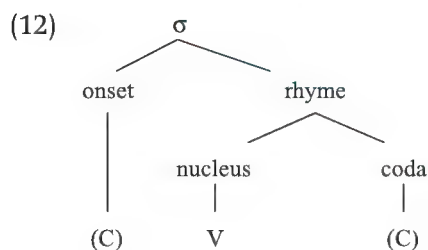
sounds [ʃ] and [ʒ] are palatal sounds, this process is called **palatalization**. You can also observe that regardless of where primary stress falls in the base, the derived form has stress on the penultimate (next to the last) syllable.

But the choice of allomorphs is not entirely predictable. For example, it's not clear if we can predict when we will get *-ation*, say, as opposed to *-ion* on a particular verb base: we find *-ation* on the verbs *accuse* and *refute*, but not in *transfuse* and *prosecute*; those have the *-ion* allomorph. The derived noun form from *combust* is *combustion*, but that of *infest* is *infestation*. Why not *combustation* and *infestation* instead? The verb base *propose* yields *proposition*, but *accuse* yields *accusation*. Why not *proposition*, or *accusation*? To some extent the choice of allomorphs seems to be quite arbitrary. We will leave this affix here, but return to it in [Section 9.5](#), where we will consider why this affix (and a number of others) display such pervasive and unpredictable allomorphy.

9.3 Other Morphology–Phonology Interactions

Not all morphology–phonology interactions have to do with allomorphy. Here we will look at two cases of reduplication in which phonological structure is important to morphological analysis. In [Chapter 5](#) we learned about full reduplication and partial reduplication. Here, we will look more closely at partial reduplication, as in some languages the description of just what in a word gets reduplicated depends on understanding the syllable structure of words.

Before we look at reduplication, though, we need to look briefly at how the sound segments of words are organized into larger units of structure. Briefly, phonologists argue that phonemes can be organized into **syllables** and syllables into larger units of structure. Syllables are composed of three parts, called the **onset**, the **nucleus**, and the **coda**, as you see illustrated in (12). The only obligatory part of a syllable is the nucleus, which typically consists of a vowel. Any consonants that go before the nucleus are called the onset, and any that go after the onset are called the coda. The nucleus and the coda form a grouping that we call the **rhyme**. The symbol that we use for 'syllable' is the Greek letter sigma.



How do we know how to divide words into syllables? This is a difficult question, but here we will use a rule-of-thumb called 'maximize the onset'. In order to find the boundaries between syllables, we can follow

the simple procedure illustrated in (13). Consider, for example, the words *opera* and *compensate* in English.⁵

- (13) a. Each vowel is the nucleus of a syllable. Mark the nucleus of each syllable

nu	nu	nu	nu	nu	nu									
o	p	e	r	a	c	o	m	p	e	n	s	a	t	e

- b. Make any consonants before the first vowel an onset. Then look at the consonants between the first and second nuclei. If there's only one, make it the onset of the following syllable. If there's more than one, put as many in the onset of the second syllable as could begin a word. Repeat the same process between the second and third, and so on. For example, the *p* and *r* in *opera* can go with the vowels that follow them, but in *compensate*, only the *p* and *s* can go with the vowels that follow them, as words in English cannot begin with the sequences *mp* or *ns*.

nu	o	nu	o	nu	o	nu	o	nu	o	nu				
o	p	e	r	a	c	o	m	p	e	n	s	a	t	e

- c. Any consonants that have not been put in the onset of a syllable go into the coda of the preceding syllable.

nu	o	nu	o	nu	o	nu	co	o	nu	co	o	nu	co	
o	p	e	r	a	c	o	m	p	e	n	s	a	t	e

- d. Join the nucleus and the coda into a rhyme. If there's no coda, just write rhyme above the nucleus.

r	r	r	r	r	r									
			Λ	Λ	Λ									
nu	o	nu	o	nu	co									
o	p	e	r	a	c	o	m	p	e	n	s	a	t	e

- e. Join the onset and the following rhyme.

σ	σ	σ	σ	σ	σ									
	Λ			Λ										
r	r	r	r	r	r									
nu	o	nu	o	nu	co									
o	p	e	r	a	c	o	m	p	e	n	s	a	t	e

This procedure can be used for English, but it can also be used for other languages as well as long as we have some idea of what consonants can begin words in that language.

5. In the examples in (13) I use English orthography rather than phonetic transcriptions, but for the beginner it sometimes helps to transcribe words.

Once we have an idea of how we can divide words into syllables, we can distinguish different types of syllables. For our purposes, it will be enough to know the difference between **heavy syllables** and **light syllables**. A heavy syllable is one that either has a long vowel or ends in a consonant (in which case it has a coda).⁶ A light syllable has a short vowel and no coda.

What does this have to do with reduplication? Interestingly, there are some languages in which partial reduplication rules require the copy to be a particular type of syllable. Take, for example, Mokilese, where the progressive form of the verb is formed by reduplicating a heavy syllable:

(14) *Mokilese (Austronesian)* (data from [Inkelas 2014: 170](#))

	<i>verb</i>	<i>progressive form</i>
'plant'	pɔdɔk	pɔd-pɔdɔk
'tear'	soorɔk	soo-soorɔk
'find'	diar	dii-diar

In the first example, the first syllable is just *pɔ*, but this is a light syllable, so the reduplication rule takes the consonant of the second syllable onset and adds it on to *pɔ* to make the heavy syllable *pɔd*. In the second example, the first syllable of the verb is already heavy (it has a long vowel), so there's no need to add anything from the second syllable to make up the copy. The third example requires another kind of adjustment to the copy for it to be heavy. In this language a sequence of two vowels can belong to different syllables. Since the second syllable of the base begins with a vowel, we can't use an onset consonant to fill out the coda of the reduplicating syllable. So the reduplication rule has to lengthen the first vowel to make a heavy syllable for the copy.

In the Pama-Nyungan language Diyari, the partial reduplication rule, which is used for a number of different purposes, copies what's called a 'minimal word', which for our purposes can be defined roughly as a unit consisting of two syllables (data from [Inkelas 2014: 170](#)):

(15) <i>Diyari</i>	wila	wila-wila	'woman'
	kanku	kanku-kanku	'boy'
	ku kuŋa	ku ku-ku kuŋa	'to jump'
	tiilparku	tiilpa-tiilparku	'bird species'

If we were to look only at the first two examples in (15), we might think that Diyari has a rule of full reduplication, but the third and fourth examples show that this is not the case; if the base word consists only of two syllables, that is what gets reduplicated, but if the base is longer, the first two syllables (the minimal word) gets reduplicated.

The reduplication rules of Mokilese and Diyari show us that knowing about the syllable structure of words can help us to understand how morphological rules work. This is true not only in the case of reduplication, but also in examples of morphology that can be found in English itself. What I have in mind here is the formation of hypocoristics. We briefly mentioned hypocoristics, the formation of nicknames, in

6. This is a slight simplification, but it will be sufficient for our purposes.

Chapter 5, noting that it was a subtractive process in English. But what we did not look at was the phonological basis of the process, which also has to do with syllable structure. Just to give you a taste of how it works, consider the challenge below.

Challenge

Consider the names and corresponding nicknames below. Can you figure out what sort of syllable the subtractive process of hypocoristic formation needs to leave behind to have a well-formed nickname in English?

Patricia	Pat or Trish
Michael	Mike
Jennifer	Jen
David	Dave
Elizabeth	Liz
Douglas	Doug

You might want to transcribe these names into IPA and divide them into syllables using the procedure in (13) to help with your analysis.

9.4 How To: Morphophonological Analysis

So far we've looked at morphophonological processes in a number of different languages. In this section, we'll take a close look at how morphologists go about analyzing allomorphy in a language that's unfamiliar. Take a look at the data in (16) from the Austronesian language Tagalog (Schachter and Otanes 1972: 290–1):

(16)	<i>Verb root</i>	<i>Actor focus or derived verb form⁷</i>	
	anak	manganak	'give birth (to)'
	bakya	mambakya	'hit with a wooden shoe'
	dukot	mandukot	'steal'
	gulo	manggulo	'create disorder'
	hiwa	manghiwa	'cut (sthg. intentionally)'
	kailangan	mangailangan	'need'
	ligaw	manligaw	'pay court to'
	manhid	mamanhid	'get numb'
	nood	manood	'watch'
	pili	mamili	'choose (several things)'
	sakit	manakit	'cause pain'
	takot	manakot	'frighten (several people)'
	walis	mangwalis	'hit with a broom'

7. The data in (16) represent two different types of verbs in Tagalog that are formed with the same prefix. The 'actor focus' verbs are roughly like active (as opposed to passive) verb forms in English, and what Schachter and Otanes call "derived" verb forms are ones that denote destructive activity or activity directed at several objects or people. For the purposes of this problem, it doesn't matter that the prefix has several different uses.

The first thing that should leap out at you when you see these data is that the forms in the right-hand column (let's refer to them as the derived forms) seem to have some sort of prefix. But it's not always exactly the same prefix. The prefix looks like *mang-* in the first form, *mam-* in the second, and *man-* in the third. There's clearly some allomorphy displayed in this set of data. To see better what's going on, it is often a good strategy to rearrange the data so that similar forms are put together. This allows you to begin to see patterns. There are a number of ways of doing this, but in (17), I've rearranged them into four groups. In the first, we can clearly segment off the prefix *mang-*. In the second group, it's still possible to segment off a prefix and leave behind something that looks exactly like the verb root. What's left over is either *mam-* or *man-*. The examples in (17c) look like something is missing, though. If we were just putting together a prefix with the stem, we might expect *mammanhid* and *mannood*, rather than the forms we actually find. And finally in (17d), we have forms in which the initial consonant of the verb root clearly seems to be absent.

(17)	a.	Verb root	Actor focus verb	
		anak ⁸	manganak	'give birth (to)'
		gulo	manggulo	'create disorder'
		hiwa	manghiwa	'cut (sthg. intentionally)'
		walis	mangwalis	'hit with a broom'
	b.	bakya	mambakya	'hit with a wooden shoe'
		dukot	mandukot	'steal'
		ligaw	manligaw	'pay court to'
	c.	manhid	mamanhid	'get numb'
		nood	manood	'watch'
	d.	pili	mamili	'choose (several things)'
		takot	manakot	'frighten (several people)'
		sakit	manakit	'cause pain'
		kailangan	mangailangan	'need'

Let's start with the forms in (17a) and (17b) that are easily segmentable. Segmenting the derived verbs, we find the allomorphs *mam-*, *man-*, and *mang-*. The first occurs on a root beginning with [b], the second with roots beginning with [d] and [l], and the third with roots beginning with [ʔ], [g], [h], or [w]. This far, the data should not surprise you: Tagalog seems to have a process of place assimilation, just as English and Zoque do. The labial-final allomorph *mam-* occurs with a labial consonant, the alveolar-final allomorph *man-* occurs with alveolar-initial roots, and the velar-final allomorph *mang-* (phonetically *maŋ-*) occurs with roots that begin with either velar or glottal consonants. So far, things seem fairly neat.

When we turn our attention to the data set in (17c), however, we will see that there's more to be said about the derived forms of the verb. It's

8. Although the spelling suggests that this form begins with a vowel, it is pronounced with a glottal stop before the vowel, so phonetically it is actually [ʔanak].

not so clear how to segment the forms in this set. Suppose we assume that the prefix is *mam-* or *man-*, as it was in (17a,b); we are then left with *anhid* or *ood* as allomorphs of the roots. Alternatively, we might assume that the prefix in these cases is just *ma-*. If we do so, then the bases would be exactly the same as the roots, namely *manhid* and *nood*. This might seem like the best solution at the moment – just adding another allomorph to the set we already have of *mang-*, *man-*, and *mam-*, but let's keep our minds open to both solutions until we've finished looking at all the data.

The data in (17d) present us with a new problem. If we assume that Tagalog has place assimilation, we would expect that we would put together *mam-* with *pili* to get *mampili* and *man-* with *takot* to form *manta-kot*, and so on. But instead we get *mamili* and *manakot*. The initial consonant of the root seems to disappear when the prefix is attached. Now, if we look back and compare the verb roots that we find in (17b) and compare them to those we find in (17d), we will see that the former begin with voiced labial or alveolar consonants, whereas those in (17d) begin with voiceless consonants. It looks like when the prefix attaches to a base that begins with a voiceless consonant, the prefix first assimilates to the place of articulation of the following voiceless consonant, and then that consonant disappears.

Here, we need to stop and think more about the assimilation rule. In Section 9.2 I suggested that each set of allomorphs has a single underlying form, from which the others are derived by phonological rule. Since place assimilation in Tagalog seems to be a predictable process, we would assume this to be the case here as well. So we need to decide at this point what the underlying form should be. Remember that it's often a good strategy to pick as the underlying form the allomorph that has the widest distribution, in other words the one that occurs with the most classes of sounds. Here, as the data in (17) show, the allomorph *mang-* occurs with glottal-initial roots ([h] and [ʔ]), as well as with velars ([g], [k], [w]). The allomorph *mam-* occurs only with labial-initial roots ([p], [b]), and the *man-* allomorph only with alveolar-initial roots ([d], [t], [l]). We can reasonably make the hypothesis then that the underlying form of the prefix is *mang-*, since it occurs with two different classes of sounds. If so, then the forms in (17) have the following underlying representations:

- (18) a. *mang* + *anak*
 mang + *gulo*
 mang + *hiwa*
 mang + *walis*
 b. *mang* + *bakya*
 mang + *duko*
 mang + *ligaw*
 c. *mang* + *manhid*
 mang + *nood*
 d. *mang* + *pili*
 mang + *takot*
 mang + *sakit*
 mang + *kailangan*

Now we can give an informal statement of two phonological rules that will derive the allomorphs from the underlying forms:

- (19) a. The nasal of *mang-* assimilates to the place of articulation of a following consonant.
 b. A voiceless consonant is deleted when preceded by a nasal consonant.

For the forms in (17a), neither rule applies. For those in (17b), only the first applies, since the verb roots don't begin with voiceless consonants. But in (17d) both rules apply.

- (20) No rules: *mang + gulo* → *mang + gulo*
 Rule (19a) only: *mang + bakya* → *mam + bakya*
 Both rules: *mang + pili* → *mam + pili* → *mam + ili*

What about the examples in (17c), however? Remember that we were undecided as to whether the allomorph of the prefix should end in a nasal at all, and we were leaning towards the solution in which the allomorph was *ma-*, as it would allow us to say that the verb roots always had the same form. We can see now, however, that that might not be the right solution. Suppose that we were to assume that the correct allomorph is *ma-*; since we have postulated that the underlying version of the prefix is *mang-*, we would have to derive *ma-* from underlying *mang-*. This in turn would require us to add a third rule that would delete *-ng* before a nasal initial verb root:

- (21) *mang + manhid* → *ma + manhid*

Before we add this third rule, however, let's see what happens if we assume that for these bases the allomorphs for the prefix are *mam-* or *man-*. Note that we already have an assimilation rule that accounts for which of the allomorphs shows up – we get *mam-* before an *m-* initial root, and *man-* before an *n-* initial root. And we already have a rule that deletes consonants after a suffix that ends in a nasal. If we tweak that rule slightly, we can derive the forms in (17c) without adding a third rule:

- (22) A nasal or voiceless consonant is deleted when preceded by a nasal consonant.

The forms in (17c) can then be derived as follows:

- (23) *mang + manhid* → *mam + manhid* → *mam + anhid*
mang + nood → *man + nood* → *man + ood*

So although it seemed at first that assuming the allomorph of the prefix in (17c) to be *ma-* made more sense, looking at the bigger picture, making the other choice allows us to derive all the allomorphs using a simpler set of rules. We assume then that this is the right solution. We must keep in mind, though, that we've only looked at a tiny set of data. If we were to continue looking at the morphology and phonology of Tagalog, we might decide that the analysis we've decided upon here needs to be revised again.

9.5 Lexical Strata

What we have seen in this chapter is that building complex words is frequently accompanied by phonological effects such as assimilation or vowel harmony. In this section we will see that in some languages such phonological effects do not apply uniformly across the entire lexicon of the language, but instead are confined to a subset of the lexicon. Indeed some languages have two or more different layers to their lexicons which behave differently in terms of phonological effects. In this section we will look at three such languages, English, Dutch, and French.

9.5.1 English

As we saw in [Section 9.2.2](#), the suffix *-tion* is associated with complex and partially unpredictable allomorphy, both of the suffix itself and of the bases it attaches to. It turns out that it's not the only suffix in English that acts that way. Consider the examples in [Table 9.2](#).

All seven of the suffixes in [Table 9.2](#) are non-native to English. Specifically, they were borrowed from Latin either directly or by way of French. All of them are like *-tion* in showing complex patterns of allomorphy. When they are added to bases, the final consonants of those bases sometimes change:

- (24) sacrifice [s] sacrific-ial [ʃ]
 Christ [t] Christ-ian [tʃ]
 dialogue [g] dialog-ic [dʒ]
 allude [d] allus-ive [s]
 historic [k] historic-ity [s]
 delude [d] delus-ory [s]
 decide [d] decis-ion [ʒ]

Table 9.2 Some non-native suffixes in English

Affix	Rule	Stem change	Stress change	Attaches to bound bases	Attaches to words	Attaches to non-native bases	Attaches to native bases
-al	N→A	sacrificial	architectural	minimal	architectural	yes	(tidal)
-ian	N→N,A	Christian	contrarian	pedestrian	Bostonian	yes	(earthian)
-ic	N→A	dialogic	Germanic	geographic	problematic	yes	no
-ive	V→A	allusive	alternative	nutritive	impressive	yes	(talkative)
-ity	A→N	historicity	historicity	atrocitiy	similarity	yes	(oddity)
-ory	V→A	delusory	excretory	perfunctory	contradictory	yes	no
-tion	V→N	decision	revelation	perception	restoration	yes	(starvation)

The stress pattern on the base often changes as well:

- (25) 'ar.ch.itec.ture ar.chi.'tec.tu.ral
 'con.tra.ry con.'tra.ri.an
 'Ger.man Ger.'ma.nic
 'al.ter.nate al.'ter.na.tive
 hi.'sto.ric hi.sto.'ri.ci.ty
 ex.'crete 'ex.cre.to.ry⁹

Furthermore, all of these suffixes can attach either to bound bases or to full words. And all of them prefer to attach to bases that are themselves non-native to English. The items in the last column in Table 9.2 are in parentheses because they are among the few native bases (sometimes the only one) on which these affixes can be found.

If we now look at suffixes that are native to English – that is, suffixes that were present in Old English, rather than borrowed from some other language – we find quite a different pattern: consider Table 9.3.

When these suffixes attach to bases, they change neither the sounds of those bases nor their stress pattern:

- (26) 'pood.le 'pood.le.dom
 'so.rrow 'so.rrow.ful
 'neigh.bor 'neigh.bor.hood
 'her.mit 'her.mit.ish
 'bo.ttom 'bo.ttom.less
 'ha.ppy 'ha.ppi.ness

Typically they attach freely to either native or non-native bases, but they do not attach to bound bases; the word *vengeful* is in parentheses because it seems to be the only example where one of these suffixes

Table 9.3 Some suffixes native to English

Affix	Rule	Stem change	Stress change	Attaches to bound bases	Attaches to words	Attaches to non-native bases	Attaches to native bases
-dom	N→N	none	none	no	kingdom	yes	yes
-er	V→N	none	none	no	writer	yes	yes
-ful	N→A	none	none	(vengeful?)	sorrowful	yes	yes
-hood	N→N	none	none	no	knighthood	yes	yes
-ish	N,A→A	none	none	no	mulish	yes	yes
-less	N→A	none	none	no	shoeless	yes	yes
-ness	A→N	none	none	no	happiness	yes	yes

9. Note that the stressing in the last pair in (25) is American English. Speakers of other dialects of English might stress these words differently.

might be said to be attached to a bound base, but it's a questionable example, since *venge*, according to the *OED*, is an obsolete word in English.

In fact, the different behavior of the two sets of affixes can be nicely illustrated by comparing the suffixes *-ic* and *-ish*, both of which can take nouns and make adjectives from them. Compare the adjectives they form from the non-native base *dialogue*:

- (27) dialogue dialogic dialoguish
 ['dɑɪ.ə.ləɡ] [dɑɪ.ə.'lɑ.dʒɪk] ['dɑɪ.ə.lə.ɡɪʃ]

The suffixes themselves differ only in their final sound, but *-ic* both changes the final consonant of its base and causes its stress pattern to change, whereas *-ish* has neither effect. What this illustrates is that English derivational morphology exhibits two different **lexical strata**, layers of lexeme formation that display different phonological behavior.

We can make one more interesting observation about the lexical strata of English. Consider the derived words in (28):

- (28) a. Two non-native suffixes: -al + -ity sequentiality
 -ian + -ity Christianity
 -tion + -al organizational
 -ive + -ity productivity
 b. Two native suffixes: -ful + -ness sorrowfulness
 -less + -ness hopelessness
 -er + -hood riderhood
 -er + -less printerless
 c. Native outside non-native -al + -ness sequentialness
 -ian + -ness Christianness
 -tion + -less organizationless
 -ive + -ness productiveness
 d. Non-native outside native -hood + -al *knighthoodal
 -ish + -ity *mulishity
 -less + -ity *shoelessness
 -ness + -ic *happinessic

Not every suffix can attach to other suffixed words in English, but sometimes we can get complex words with two or more layers of suffixes. As (28) shows, we can often affix a non-native suffix to a base that already has a non-native suffix, and similarly put a native suffix on a base that already has a native suffix. Further, we can often stack up two suffixes if the first (the innermost in terms of structure) is non-native and the second native. What's much more difficult – although not absolutely impossible, as we will see in [Chapter 10](#) – is to first affix a native suffix and then put a non-native suffix outside it. This makes perfect sense: non-native suffixes prefer to attach to non-native bases. Once a native suffix has been added to a base, regardless of whether that base was native or non-native to begin with, the derived word counts as a native word as far as further affixation is concerned.

The suffixes we've looked at here show rather different behavior, which suggests that English derivational morphology displays two different lexical strata. To be honest, not all suffixes in English are as easily

classified as the ones we've looked at in this section. While other suffixes native to English behave much as those discussed here, this is not the case with all non-native suffixes. Some suffixes that are borrowed, and therefore should be part of the non-native stratum of English, behave more like native suffixes in that they have no phonological effects on their bases and attach indiscriminately to both native and non-native bases. We will not go further into the intricacies of English derivation here, but merely point out that while there is a clear suggestion that English suffixes belong to two different strata, there is some blurring between them. Further, were we to look at prefixation rather than suffixation, the lines between strata would be even more blurred.

9.5.2 Dutch and French

Dutch and English are closely related languages, and they share a history of contact with French and Latin. It is therefore not surprising that the morphology of Dutch exhibits two lexical strata, just as English does. We'll give just a brief illustration here. In Dutch, the suffix *-heid* 'ness' is of native origin, and *-iteit* 'ity' is non-native. As we saw in English, the native suffix attaches easily either to native or non-native bases, but the non-native one can only occur on non-native bases (Booij 2002: 95):

- | | | | | |
|------|---------------|-------------|---------------|-------------------|
| (29) | <i>-heid</i> | blindheid | 'blindness' | (native base) |
| | | diversheid | 'diverseness' | (non-native base) |
| | <i>-iteit</i> | *blinditeit | | (native base) |
| | | diversiteit | 'diversity' | (non-native base) |

As in English, non-native affixes can occur on either bound bases or free words, whereas native affixes only occur on free words. And as in English, if a word contains both native and non-native affixes, the native ones must occur outside the non-native ones.

What may be somewhat more surprising is that French, a language itself descended directly from Latin, also shows signs of lexical strata (Huot 2005). French suffixes can be divided into those that are called 'popular' (in French 'populaire') and those that are called 'learned' (in French 'savant'). The former have descended from Latin undergoing all the sound changes that the vocabulary of French has been subject to. The latter come from scholarly Latin by borrowing later in the history of French. Popular suffixes typically attach to popular roots, and learned suffixes to learned roots (Huot 2005: 65):

- (30) *Popular suffixes that prefer popular roots*
- | | |
|------|---|
| -age | doublage 'doubling', grattage 'scratching' |
| -ier | jardinière 'gardener', pétrolier 'oil-tanker' |
| -eux | chanceux 'lucky', venteux 'windy' |
- Learned suffixes that prefer learned roots*
- | | |
|-------|---|
| -ion | inscription 'inscription', punition 'punishment' |
| -if | actif 'active', duratif 'durative' |
| -aire | articulaire 'articular', réfractaire 'refractory' |

Popular suffixes sometimes do attach to learned roots, but learned suffixes do not attach to popular roots:

- (31) *Popular suffix on learned root*
infectieux ‘infectious’, *torrentueux* ‘torrential’

Popular suffixes tend not to attach to already suffixed words, but learned suffixes can sometimes attach to other learned suffixes:

- (32) *démiss* + *ion* + *aire* *démissionnaire* ‘one who has resigned’

And finally, popular roots sometimes have corresponding learned allomorphs:

- | (33) Popular | | Learned | |
|--------------|----------|---------------|------------|
| angle | ‘angle’ | angul + aire | ‘angular’ |
| cercle | ‘circle’ | circul + aire | ‘circular’ |
| peuple | ‘people’ | popul + aire | ‘popular’ |

So what we see here is that there are two sets of affixes that display somewhat different patterns of behavior. The lexicon of French thus gives us another example where morphology is not neat and homogeneous, but instead seems to be organized into two relatively discrete layers.

Summary

In this chapter we have looked at the connection between phonology and morphology. Morphemes frequently have allomorphs, phonologically distinct variants that occur in different environments. Sometimes, as we saw, those environments are predictable, and we can postulate phonological rules that explain the distribution of the allomorphs. Indeed, we can often postulate a single underlying phonological form from which all the allomorphs can be derived. We have looked at a number of typical kinds of phonological rules that explain allomorphy in various languages: assimilation of various sorts, epenthesis, vowel harmony, and intervocalic voicing. We have also seen that not all allomorphy is entirely predictable; as the morphology of English shows, it can be quite unpredictable where one allomorph or another shows up. And we have looked at morphophonological processes that depend on syllable structure. Finally, we have looked at three cases in which different segments of the lexicon constitute different lexical strata displaying different phonological behavior or different patterns of allomorphy.

Exercises

- The following forms are from Wappo (Yukian) (Thompson, Park, and Li 2006: 125–7). The first set of examples is glossed for you. Using them

as a model, first analyze the next two sets of data and then answer the questions below.

- | | | |
|----|---------------------|---------------------------|
| a. | olol-asa? | 'is making X dance' |
| | dance-CAUS:DUR | |
| | olol-is-ta? | 'made X dance' |
| | dance-CAUS-PST | |
| | olol-is-ya:mi? | 'will make X dance' |
| | dance-CAUS-FUT I | |
| | olol-asi? | 'make X dance!' |
| | dance-CAUS: IMP | |
| | olol-asa-lahkhî? | 'isn't making X dance' |
| | dance-CAUS: DUR-NEG | |
| | olol-is-ta-lahkhî? | 'wasn't making X dance' |
| | dance-CAUS-PST-NEG | |
| | olol-is-lahkhî? | 'don't make X dance!' |
| | dance-CAUS: IMP-NEG | |
| b. | hicasa? | 'is making X pound Y' |
| | hicista? | 'made X pound Y' |
| | hiciya:mi? | 'will make X pound Y' |
| | hicasî? | 'make X pound Y!' |
| | hicasalahkhî? | 'isn't making X pound Y' |
| | hicistalahkhî? | 'wasn't making X pound Y' |
| | hicialahkhî? | 'don't make X pound Y' |
| c. | hintô?asa? | 'is making X sleep' |
| | hintô?ista? | 'made X sleep' |
| | hintô?isya:mi? | 'will make X sleep' |
| | hintô?asi? | 'make X sleep!' |
| | hintô?istalahkhî? | 'wasn't making X sleep' |
| | hintô?islahkhî? | 'don't make X sleep!' |

Once you have segmented and glossed the (b) and (c) sets of data, list all the allomorphs of all morphemes. Is all of the allomorphy predictable? Where you can, explain informally what seems to determine the distribution of the allomorphs.

2. In Tagalog, the circumfix *ka ... an* creates nouns designating the class or group of whatever the base denotes. Identify all allomorphs in the forms below and explain their distribution using informal phonological rules (Schachter and Otones 1972: 101):

banal	'devout'	kabanaan	'devoutness'
bukid	'field'	kabukiran	'fields'
bundok	'mountain'	kabundukan	'mountains'
lungkot	'sadness'	kalungkutan	'sadness'
pangit	'ugly'	kapangitan	'ugliness'
pulo	'island'	kapuluan	'archipelago'
dagat	'sea'	karagatan	'seas'
dalita	'poverty'	karalitaan	'poverty'
Tagalog	'a Tagalog'	katagalugan	'the Tagalogs'

3. The Dutch diminutive has several allomorphs. Determine what they are, and explain their distribution (De Haas and Trommelen 1993: 279):

a.	gum	'eraser'	gumetje
b.	roman	'novel'	romanetje
c.	parasol	'parasol'	parasoletje
d.	kar	'cart'	karetje
e.	lichaam	'body'	lichaampje
f.	pruim	'plum'	pruimpje
g.	bezem	'broom'	bezempje
h.	koning	'king'	koningkje
i.	haring	'herring'	haringkje
j.	streep	'stripe'	streepje
k.	kabinet	'cabinet'	kabinetje
l.	almanak	'almanac'	almanakje
m.	wereld	'world'	wereldje
n.	banaan	'banana'	banaantje
o.	tuin	'garden'	tuintje
p.	kuil	'hole'	kuiltje
q.	altaar	'altar'	altaartje

HINT: Examples a–d have short vowels in their final syllables. Examples n–q have long vowels or diphthongs in their last syllables.

4. Form the plurals of the following words in English, and transcribe them in the IPA:

lip	lathe
pot	kiss
tack	buzz
club	church
thud	garage
thug	judge
cliff	arena
path	hero
stove	

- How many allomorphs are there for the plural morpheme in English?
 - Which of the allomorphs makes the best candidate for the underlying form of the plural morpheme?
 - Formulate a phonological rule that derives the various allomorphs of the plural morpheme from the underlying form.
5. Thinking about the pattern you discovered in exercise 4, now consider the plurals of the following words:

wolf
calf
house
mouth
elf
knife

How do these differ from the plurals you discussed above?

6. In exercise 1 of Chapter 5 you looked at a process of infixation in the Austronesian language Leti. Some of the data you looked at there are given again in (a) (Blevins 1999):

a.	kakri	'cry'	kniakri	'the act of crying'
	pali	'float'	pniali	'the act of floating'
	sai	'climb'	sniai	'the act of climbing'
	teti	'chop'	tnieti	'the act of chopping'
	vaka	'ask'	vniaka	'the act of asking'
	vanunsu	'knead'	vnianunsu	'massage' = 'the act of kneading'

Now compare those data to the ones in (b):

b.	kili	'look'	knili	'the act of looking'
	kini	'kiss'	knini	'the act of kissing, kiss'
	surta	'write'	snurta	'the act of writing, memory'
	tutu	'support'	tnutu	'the act of supporting, support'
	virna	'peel'	vnirna	'the act of peeling'

What are the two allomorphs of the nominalizing affix in Leti? What determines which allomorph goes with which bases?

7. Consider the data below from the Mayan language Tzutujil (Dayley 1985: 206–7):

K'uluuj	'to meet, encounter'	k'ulaani	'married'
jaqooj	'to open'	jaqali	'open'
d'eb'ooj	'to stain with thick liquid'	d'eb'eli	'thick (of liquid)'
b'olooj	'to twine, boil meat'	b'olaani	'cylindrical'
d'oyooj	'to cut with an axe'	d'oyoli	'cuttable'
wonooj	'to push with the head'	wonoli	'bent over'
ketooj	'to cut with a very sharp machete'	keteli	'discoid, wheel-shaped'
ch'ikooj	'to clean land for tilling'	ch'ikili	'stuck in'
jotooj	'to raise'	jotoli	'be above'
ch'anooj	'to spank a naked person'	ch'anali	'naked'

Identify all allomorphs and try to state the conditions under which each occurs. What morphophonological process is illustrated by the data in the second column?

8. The following data are from the constructed language (conlang) Dothraki that figures in the HBO TV series *Game of Thrones*. Consider the forms below. Segment the forms in the second column into a base and a suffix and give a meaning for the suffix. What are the allomorphs of this suffix? Propose an underlying form and phonological rules to derive the surface forms from the underlying form (data from http://wiki.dothraki.org/Derivational_morphology).

fonak	'hunter'	fonakasar	'hunting party'
oqet	'sheep'	oqeteser	'flock of sheep'
zir	'bird'	zirisir	'flock of birds'
jano	'dog'	janosor	'pack of dogs'

9. Consider the data below from Hopi, a Uto-Aztecan language spoken in the southwestern part of the United States (data from [Jeanne 1982: 246–8](#)). Describe the allomorphy that is exhibited in the future tense in Hopi. Propose an underlying form for each verb and state the phonological rules that can be used to derive the surface forms from the underlying forms.

<i>non-future</i>	<i>future</i>	
ʔiɪya	ʔiɪni	'plant'
nöösa	nösni	'eat'
wiɪvi	wipni	'climb'
piɪwi	piwni	'sleep'
qaaci	qacni	'be in position (inanimate)'
siɪwi	siwni	'dizzy'
caama	camni	'take out'
heeva	hepni	'seek'

10. Pame is an Otomanguean language spoken in Mexico. Identify the singular/dual and plural affixes in Pame. If you observe allomorphy of either affix or base, describe the allomorphy, and propose underlying forms and phonological rules deriving the surface form from the underlying form (data from [Gibson and Bartholemew 1979: 310](#)).

<i>sg or dual</i>	<i>pl</i>	
ŋgobéʔet	mbéʔet	'flag'
ŋgodèocʔ	ndèocʔ	'bridge'
ŋgokhwèʔ	ŋkhwèʔ	'bean'
ŋgosáon	nsáon	'night'
ŋgolhwá	nlhwá	'ear of corn'
ŋgopʔóho	mbʔóho	'seat'
ŋgokwáŋ	ŋgwáŋ	'tree, piece of wood'
ŋgohwéʔ	hwéʔ	'thorn'
ŋgomhé	mhé	'tortilla'

10 Theoretical Challenges

CHAPTER OUTLINE

KEY TERMS

Item and Arrangement

Item and Process

Word and Paradigm

realizational model

multiple exponence

Lexical Integrity Hypothesis

blocking

rival affixes

competition

bracketing paradoxes

In this chapter you will get your first taste of the theoretical challenges that morphologists face.

- ◆ We will consider what the best way is to characterize morphological rules.
- ◆ And we will examine several theoretical controversies that have occupied morphologists in recent years: the extent to which rules of morphology and rules of syntax can interact, the nature of blocking, the best way to analyze so-called bracketing paradoxes, and the characterization of affixal polysemy.

10.1 Introduction

Up to this point, we've spent a lot of time looking at the way morphology works in languages – what kinds of morphemes there are, what to call them, how to analyze data, and so on. This is an important and necessary first step to becoming a morphologist, but there's more to morphology than just being able to analyze data. When we study morphology, or indeed any of the other subfields of linguistics, we have a much larger goal in mind, which is to characterize and understand the human language faculty. Put a bit differently, our ultimate goal as linguists is to figure out the way in which language is encoded in the human mind. For morphologists, our specific goal is to figure out how the mental lexicon is encoded in the mind. Doing so requires us to model the mental representation of language, to make claims about exactly what morphological rules look like, and to propose hypotheses about what is possible in human languages and what is impossible.

A good hypothesis about language is one which is empirically testable: it should be clear what sort of data to look for that would disprove the hypothesis. To illustrate this, let's look first at two examples of theoretical hypotheses that have been proposed by morphologists. I start with these two precisely because they make clear claims about what sorts of morphology we should expect to find in languages and because these claims have subsequently been disproven.

The first hypothesis we'll look at is called the **Righthand Head Rule**:

- (1) *The Righthand Head Rule* (Williams 1981: 248)

"In morphology, we define the head of a morphologically complex word to be the righthand member of that word."

You'll recall from [Chapter 3](#) that the head of a compound was the morpheme that determined the syntactic and semantic category of the compound. Clearly, in English, it's the righthand element in the compound that's the head (so *sky blue* is syntactically an adjective like *blue* and semantically a type of blue as well). More broadly, the head of a word is that morpheme that determines the category of the word, and in languages that have gender in nouns, or inflectional classes in nouns and verbs, the head determines the gender or class of the word as well. For example, in German, the suffix *-heit* attaches to adjectives to form nouns, specifically feminine nouns. Since *-heit* determines the category and gender of the derived noun, it is the head of the word. The Righthand Head Rule is a theoretical hypothesis that basically says that all compounds should be right-headed, and only suffixes (and not prefixes) can be the heads of words.

At first glance, this hypothesis is plausible enough when we look at English. Compounds are indeed right-headed, and for the most part in English it's the suffix that determines the category of a complex word. However, the Righthand Head Rule can easily be disproven by looking at

data from other languages. For example, in [Chapter 3](#) we saw that both Vietnamese and French have left-headed compounds:

- (2) *French* timbre poste ‘stamp-post’ = ‘postage stamp’
 Vietnamese nhá thuong ‘establishment be-wounded’ = ‘hospital’

And it is not hard to find languages in which prefixes change the category of words and determine their gender or class. Even English has at least one prefix that changes category, and therefore would have to be recognized as the head of the derived word:

- (3) *de-*
 debug
 delouse
 de-ice

The prefix *de-* attaches to nouns and makes verbs. Similarly, Swahili has a prefix *ku-* that forms nouns from verbs:

- (4) From [Vitale \(1981: 10\)](#)
 ku-tafutwa kwa Juma
 -ing-search for Juma
 ‘the searching for Juma’

Since this prefix determines the category of the derived word, we would have to consider it to be the head of the word. These examples show, then, that the Righthand Head Rule cannot be correct.

A second theoretical proposal that turns out not to be correct is the **Unitary Base Hypothesis**:

- (5) *The Unitary Base Hypothesis* ([Aronoff 1976: 48](#))
 “We will assume that the syntacticosemantic specification of the base, though it may be more or less complex, is always unique.
 A WFR [word formation rule] will never operate on either this or that.”

The Unitary Base Hypothesis in effect says that we should never expect to find in a language a morpheme that attaches to bases of two different categories, say adjective and noun, or noun and verb. We have seen, however, that there are many affixes that can attach to more than one base: *-ize* in English attaches to both adjectives (*legalize*) and nouns (*unionize*) to form verbs, and *-er* attaches to both verbs (*writer*) and nouns (*villager*) to form nouns.¹ It seems that affixes sometimes (in fact frequently!) do attach to “either this or that.” The Unitary Base Hypothesis makes a clear claim about what we should expect to find in the languages of the world, but that is not in fact what we find.

1. Aronoff is aware of examples like these, and is forced to argue that there are two different *-ize* suffixes and two different *-er* suffixes that are homophonous, that is, that sound identical. It is generally accepted, though, that this is not a strong defense of the Unitary Base Hypothesis, and that the hypothesis is therefore almost certainly incorrect.

Why start out a chapter on theory with two incorrect hypotheses? What is important is that these hypotheses are testable: we know what sort of data to look for, and having looked for those data can determine that these hypotheses cannot correctly characterize our theoretical model of the mental lexicon. Notice that I haven't talked about proving hypotheses to be true. In fact, it is never possible to prove a scientific hypothesis to be true: since there will always be some linguistic data we have not yet looked at, there is always some chance that further study will prove a hypothesis to be false. We take a theoretical hypothesis to be sound just as long as we have not yet found evidence against it.

Good hypotheses must therefore be testable, and we would like them to explain a wide range of data. But there is one more thing we would want of a good theoretical hypothesis. Since generative linguists are ultimately concerned with the mental representation of language, part of their concern is to explain how children acquire knowledge of those mental representations so quickly and with such ease. Our theory should help us to explain how children acquire morphological knowledge. With this in mind, we can now go on to look at a number of other theoretical proposals for characterizing our mental representation of morphology that are less easy to dismiss. Keep in mind that we can look at only a few interesting points of morphological theory here; in the last three decades there has been a great deal written about morphological theory that we will not be able to cover in this chapter.

10.2 The Nature of Morphological Rules

Up to this point we've talked about morphological rules for affixation, compounding, internal stem change, and other means of creating new words, but we have only characterized those rules informally. One of the important parts of modeling the mental lexicon is to characterize morphological rules formally. In this section we will look at different formal systems for characterizing morphological rules and try to see how they make different claims about the sorts of morphology we ought to find in the languages of the world.

10.2.1 Morphemes as Lexical Items: Item and Arrangement Morphology

Let's take another look at one of the informal rules of word formation that we proposed in [Chapter 3](#):

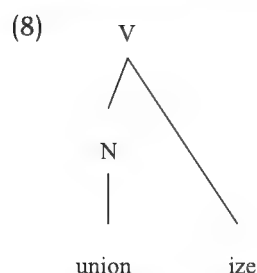
- (6) *-ize* attaches to adjectives or nouns of two or more syllables where the final syllable does not bear primary stress. For a base 'X' it produces verbs that mean 'make/put into X'.

One way of making this sort of rule formal is to assume that in our mental lexicons the morpheme *-ize* has a lexical entry, just as free morphemes do, and that part of its lexical entry is the following:

(7) *The -ize rule (more formal version)*

-ize structural information: $[[]_{A,N}]_V$
 semantic information: 'make A; make/put into N'
 phonological information: $[... \sigma \sigma_W aɪz]$

The first line of this rule gives structural information: it says that *-ize* is a suffix that attaches to nouns or adjectives, and produces verbs. In fact, it says somewhat more than this, as the brackets indicate that when the suffix is added, a bit of hierarchical structure is formed:



The second line of the rule tells us what the resulting word means; this part of the rule can be formalized as well, using special notation, but we will not do so here. We'll merely say that when the piece *-ize* is added to a base, it also adds the meaning 'make A or make/put into N'. Finally, the third line uses the Greek letter sigma (σ) to stand for 'syllable', and the subscript W to stand for a 'weak' or unstressed syllable. This, then, encodes the information that *-ize* requires a base that has at least two syllables, the last of which must not bear stress.

This kind of theory in effect makes a claim that affixes are just like free morphemes in that they have lexical entries that include various types of information. The only difference between the entry for an affix and for a base is that the affix is a bound morpheme, and therefore as part of its structural information requires another category to attach to. Theories that propose rules of this sort are traditionally referred to as **Item and Arrangement** (IA) theories, because they claim that morphemes have independent existence in the mental lexicon with their own structural, semantic, and phonological information, and that they can be arranged hierarchically into words. In a way, such theories make the claim that complex words have a sort of internal syntactic structure.

10.2.2 Morphemes as Processes and Realizational Morphology

This is, of course, not the only way of formalizing morphological rules. Indeed, many morphologists believe that it is a mistake to count morphemes as 'things' that have their own independent existence in our mental lexicons and to treat morphology as the internal syntax of words. The alternative, they argue, is to allow only free morphemes to have lexical entries of their own, and to introduce bound morphemes using rules that contain the phonological form and the semantic content of the bound morpheme. The *-ize* rule might look like (9) in such a theoretical framework:

(9) *The -ize rule*

$X \rightarrow Xize$, where $X = N, A$ and $Xize = V$ meaning ‘make, put into X ’²

In the case of the *-ize* rule, this may not look terribly different than the lexical entry we proposed above. But consider the sort of rule we would have to propose for an irregular past tense form like *sang* in English. Since the past tense in this form is created by a process of ablaut, we can propose a rule that changes the vowel [ɪ] to [æ] to produce *sang*.

(10) *Irregular past rule for the sing, swim, ring class of verbs*

$CɪN \rightarrow CæN$
 [–past] [+past]

If C stands for any consonant, and N for any nasal consonant, we can express the vowel change that takes place in the past tense very simply with such a rule. It is not so easy to see how to express an internal vowel change using the ‘morpheme as thing’ model; it doesn’t seem to make sense to say that the past tense in such verbs is a morpheme that consists of only the vowel [æ]. The sort of theoretical framework that treats morphemes as parts of rules is sometimes called an **Item and Process** (IP) theory, because morphological rules are conceived as operations or processes that act on free morphemes.

Related to the Item and Process model is the **Word and Paradigm (WP) model**. WP models are also sometimes known as **realizational models**.³ They are often proposed to account for inflectional word formation in languages that have complex paradigms, especially the sort of paradigm which exhibits a characteristic called **multiple exponence**. Multiple exponence occurs when particular inflectional characteristics – say past tense, or third person – are signaled by more than one morpheme in a word. For example, in the Latin second person singular past tense verb form *amāvisti* ‘you-sg loved’ we might say that the root is *am* and the stem with theme vowel *amā*. The past tense is signaled by the morpheme *-v*, and the second person singular morpheme is *-isti*. But this is not quite correct, because the morpheme *-isti* is a person/number ending that is only used in the past tense; in some sense, *-isti* bears the meaning of past tense along with its person/number meaning. The past tense meaning is signaled twice in the word *amāvisti*. This is what we mean by multiple exponence.

In an IA theory of morphology, multiple exponence is problematic. IA models work best when there is a one-to-one correspondence between morphemes and meanings. In other words, each morpheme expresses one and only one inflectional or derivational meaning. WP or realizational models, on the other hand, do not separate out morphemes into discrete pieces, but rather state rules that associate meanings (single or multiple) with complex forms. For example, a realizational rule for the second person singular past tense form of the verb in Latin might be (11):

2. For the moment, we can set aside how the phonological information is stated in this sort of theory.

3. This is a bit of a simplification. Although WP is realizational, other kinds of morphological theory are realizational as well. See Chapter 11 for two examples.

$$(11) \left[\begin{array}{c} X \\ +\text{past} \\ 2\text{nd person} \\ \text{sg.} \end{array} \right] \rightarrow [\text{Xvisti}]$$

The realizational model does not recognize *-v* and *-isti* as separate morphemes. Instead, the form *amāvisti* is conceived as a unit that expresses past tense, second person, and singular number conjointly. The existence of multiple exponence causes no problems in a realizational model, because inflected words need not be segmented into discrete pieces.

10.2.3 Can We Decide between Them?

You might at this point be wondering if it really makes a difference whether we conceive of morphemes as things with discrete meanings and their own lexical entries, or as parts of morphological rules that realize unanalyzable words with complex meanings. Our conception of what morphological rules look like really does matter, because it makes predictions about what sorts of morphology we should expect to find in the languages of the world. On the one hand, IA theories predict that morphology ideally should be agglutinative, with words segmentable into several pieces, each of which has a distinct meaning. In some languages this is the case – recall our sketch of Turkish in [Chapter 7](#). On the other hand, WP theories, although they do not strictly preclude agglutinative morphology, lead us to expect that morphology typically ought not to be agglutinative; rather, it should contain lots of multiple exponence, as is the case in Latin, but not in Turkish. The problem, of course, is that neither of these predictions is quite right. Some languages are more agglutinative than others. Some languages have lots of multiple exponence, and others have little or none.

So it does make a difference which kind of theory we choose. But the choice is made difficult by the complexity and variety of morphology we actually do find in the languages of the world. In addition, nothing rules out the possibility that both models might have merits. For example, it is possible that IA theories do a better job of modeling derivational morphology in many languages and IP or WP models a better job of modeling inflectional morphology (see [Borer 2013](#) for an argument to this effect).

The choice between models is made even more difficult by the fact that each of these models is not really just a single theory, but a kind of umbrella that encompasses a number of different theories. For example, in their simplest form IA models say that the correspondence between meanings and ‘pieces’ ought to be one-to-one, but IA models need not say this. It is possible to propose an IA model in which the relationship between pieces and meanings is ideally, but not strictly, one-to-one. Similarly, there are a number of different versions of WP or realizational models that are plausible. It seems safe to say, though, that each theory must be tested against a wide variety of languages with all different sorts of word formation, and a wide variety of inflectional and derivational

word formation processes including not only affixation, but also compounding, conversion, reduplication, and templatic morphology.

Challenge

Review the examples of reduplication that we discussed in [Chapter 5](#). Which of the three models discussed above (IA, IP, or WP) is best suited to modeling rules of reduplication?

For students interested in studying questions of theoretical models in more detail, a good place to start is [Spencer \(1991\)](#). Early generative theories that follow the IA model are [Lieber \(1980\)](#), [Selkirk \(1982\)](#), [DiSciullo and Williams \(1987\)](#), and [Lieber \(1992\)](#). Generative theories that take an IP, WP, or generally realizational perspective are [Aronoff \(1976\)](#), [Anderson \(1982, 1992\)](#), and [Stump \(2001\)](#). We will look in more detail at several current theoretical models in [Chapter 11](#).

10.3 Lexical Integrity

In [Chapter 8](#) we considered the relationship between morphology and syntax and saw several ways in which these two segments of the grammar are closely intertwined. The relationship between morphology and syntax indeed has given rise to one of the most interesting and longest-standing theoretical controversies among morphologists. Early in the history of generative morphology, several theorists proposed what has come to be called the **Lexical Integrity Hypothesis**. One version of this hypothesis is (12):

(12) *Lexical Integrity Hypothesis*

“No syntactic rule can refer to elements of morphological structure.” ([Lapointe 1980: 8](#))

What this means is that syntactic rules – phrase structure or movement rules, for example – cannot look into words and manipulate their internal structures. The Lexical Integrity Hypothesis requires that morphological and syntactic rules be fundamentally different. Morphological rules are concerned with affixes and bases, with rules of reduplication and ablaut, and so on, that affect the internal structure of words. Rules of syntax take words as unanalyzable wholes and form them into phrases and sentences. One way of ensuring this separation between morphology and syntax that was proposed early in the history of generative morphology was to order morphological rules before syntactic rules, as if there were something like a linguistic assembly line that started with the smallest units of structure and proceeded to larger and larger units:

The model in [Figure 10.1](#) ensures the separation of morphology and syntax because syntax only gets to look at already-formed words. Many linguists no longer believe that rules operate in a strict ‘assembly-line’ fashion, but nevertheless continue to maintain that morphological and syntactic rules must be kept separate from one another.

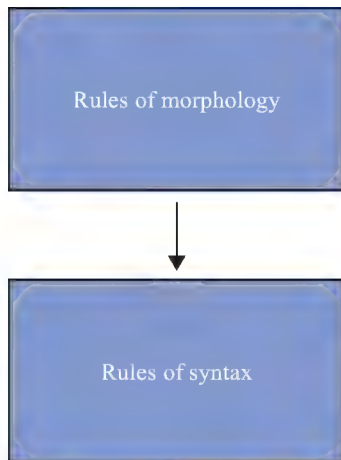


FIGURE 10.1
One way of organizing the grammar

The Lexical Integrity Hypothesis is both plausible and testable: indeed it makes a clear prediction that we should never find fully formed phrases or sentences inside words. If phrases are formed by syntactic rules and syntactic rules are separate from morphological rules, words should not contain phrases or sentences. However, as we saw in [Chapter 8](#), there are some sorts of words that do seem to contain phrases and even sentences. Among these are phrasal compounds like those in (13):

- (13) stuff-blowing-up effects
 bikini-girls-in-trouble genre
 comic-book-and-science-fiction fans
 God-is-dead theology

It is also possible in some languages – including English – to attach prefixes or suffixes to whole phrases or even whole sentences (see [Chapter 8](#) for more examples):

- (14) a. *English*
 mouse- and rat-like
 pre- and post-war
 I-can-do-it-too-ness
 b. *Turkish* ([Lewis 1967: 41](#))
 Tebrik ve teşekkür-ler-im-i
 [congratulation and thank]-PL-MY-ACC
 ‘my congratulations and thanks’

As [Bauer, Lieber, and Plag \(2013\)](#) show, in English quite a few affixes can attach to phrases and sentences and a few prefixes (*pre-*, *post-*, *hyper-*, *hypo-*) can themselves be conjoined and attached to a base. Turkish allows some of its inflectional endings to apply equally to two conjoined bases, as suggested by the bracketing in the gloss of (14b). If syntactic phrases and sentences can act as the bases for complex words, then the strict separation of morphological and syntactic rules required by the Lexical Integrity Hypothesis cannot be correct.

Linguists have therefore proposed alternatives to the Lexical Integrity Hypothesis. Some linguists have proposed models in which limited

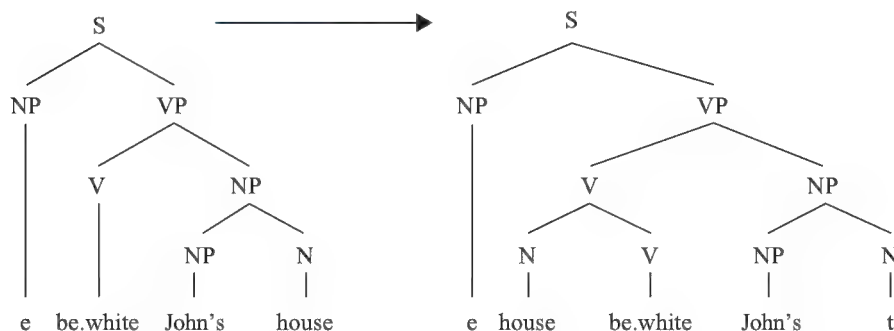
interaction between morphology and syntax is possible. Others, however, have taken a more radical approach, arguing that there should be no separation between morphology and syntax, and that syntactic rules should be responsible for at least some sorts of word formation.

One sort of word formation that has been used to argue for the hypothesis that morphology and syntax cannot be separated is noun incorporation, which we looked at briefly in [Chapter 8](#). To refresh your memory, consider the Mohawk examples in (15) ([Baker 1988: 20](#)):

- (15) a. Ka-rakv ne sawatis hrao-nuhs-a?
 3_N-be.white DET John 3_M-house-SUF
 'John's house is white'
 b. Hrao-nuhs-rakv ne sawatis
 3_M-house-be.white DET John
 'John's house is white'

In (15a) the noun 'house' is part of an independent noun phrase (NP). In (15b), however, it has been incorporated into the verb 'be white' so that together they form a single word. In a syntactic analysis of noun incorporation (15b) starts out with the noun 'house' part of an NP with 'John's'. However, a syntactic movement rule plucks 'house' from its NP and attaches it to the verb 'be white', as (16) illustrates:⁴

- (16) From [Baker \(1988: 20\)](#)



There is a great deal that might be said about the pros and cons of this analysis, although we cannot do so here. I should point out, though, that while some linguists find the evidence for this analysis convincing, others are less convinced and prefer to work within theoretical models that treat noun incorporation as the result of morphological rules, and allow less interaction between morphology and syntax. As with many other theoretical issues in morphology, the jury is still out on the best way to treat the relationship between morphology and syntax.

The interested student can read further on this topic. A general article on the status of the Lexical Integrity Principle is [Lieber and Scalise \(2007\)](#). For the debate on the analysis of noun incorporation, you might want to compare the arguments in [Baker \(1988, 1996\)](#) to those in [Mithun](#)

4. In (16) 'e' indicates that the subject NP is empty, and 't' stands for 'trace', which marks the place from which a constituent has been moved.

(1999). A good general overview of the topic of noun incorporation can be found in Massam (2009).

10.4 Blocking, Competition, and Affix Rivalry

Consider the data in (17):

- (17) a. curious curiosity
 generous generosity
 impetuous impetuosity
- b. glorious ?gloriosity
 furious ?furiosity
 gracious ?graciosity

Generally, as the examples in (17a) suggest, it seems possible in English to derive an *-ity* noun from an adjective that ends in the suffix *-ous*. But in some cases – for example, those in (17b) – the *-ity* form seems very odd, if not downright impossible. Why should this be? The examples in (17) illustrate a phenomenon that morphologists call **blocking**. In its simplest form blocking is said to occur when there is a simplex word that bears the same meaning or fulfills the same function as the purportedly non-existent derived word. In the case of the examples in (17b), but not (17a), there are simple, underived nouns from which the *-ous* adjectives are formed: *glory*, *fury*, and *grace*. Some morphologists hypothesize that since those words exist in English, they block or preclude the formation of the *-ity* nouns. The *-ous* adjectives in (17a) do not derive from free morphemes (they are built on bound bases), and therefore can form nouns by affixation of *-ity*.

We might also wonder whether blocking occurs in the formation of nominalized verbs in English:

- (18) ?occuration occurrence ?occurrent
 reservation ?reservance ?reservment
 ?amusement ?amusement amusement

As the examples in (18) suggest, there are a number of different suffixes that form nouns from verbs, among them *-ation*, *-ance*, *-ure*, and *-ment*. Sets of affixes like these that have the same function or meaning are sometimes referred to as **rival affixes**. Rival affixes are said to be in **competition** with each other, the idea being that for each verbal base, only one affix is possible. Other nominalizing affixes cannot attach, or at least they sound very odd to the ears of a native speaker. The idea is that the existence of one nominalization blocks or precludes the existence of others.

Blocking is also said to occur in inflectional paradigms as well as in cases of derivation. For example, in English the suppletive past tense form *went* is said to block the formation of a regular past tense form **goed*, and the existence of the irregular plural form *children* is said to block the formation of the regular plural **childs*.

The obvious question that a theorist might ask is why blocking should occur. One possible answer is that languages tend to avoid synonymy. Once we have a word that means ‘more than one child’ or ‘the process or result of reserving’, why would we need another? But is it true that languages always avoid synonymy? Some evidence that seems to support this hypothesis comes from the ‘doublets’ or ‘triplets’ like the ones in (19) – bases that do take more than one plural or nominalizing affix. Consider the examples in (19):

- (19) a. brothers brethren
 b. commission committal commitment

Superficially, these examples seem to provide evidence against blocking, but actually they don’t. In (19a) we have two plurals of the word *brother*, one the regular plural *brothers* and the other an archaic plural *brethren*. In (19b) we can see that the verb *commit* has three different nominalizations, not just one. But these examples do not really argue against blocking, because we don’t in fact have cases of perfectly synonymous forms. Although *brethren* was at one time just a plural of *brother*, it has specialized in meaning and is used in religious contexts to refer to members of the same church. In the case of the different nominalizations of the verb *commit*, each one has a specific lexicalized meaning. *Commission*, for example, is the act of committing, and a *commission* an order to create a piece of art. A *committal* is an order to send someone to prison or to the hospital. And a *commitment* is a pledge of certain sorts. Blocking doesn’t occur in these cases because we do not have words that are synonymous. But perhaps we shouldn’t be too hasty to draw conclusions here.

Challenge

Is it really true that languages avoid synonymy? Try to think of examples of words that you might consider to be perfectly synonymous. You may consider simplex as well as complex words.

The answer to my challenge is that it’s not all that clear that synonymy avoidance is anything more than a tendency. For example, in English we have simplex items like *couch* and *sofa* that for all intents and purposes tend to be synonyms, as are the compounds *groundhog* and *woodchuck* in North American English. With respect to derivation and inflection, at least in English, it turns out that there is evidence that synonymy is not always avoided, and indeed that blocking need not occur. Consider the examples in (20):

- (20) purity pureness
 generosity generousness
 impetuosity impetuousness

As the examples in (20) show, alongside a noun formed with *-ity* it is always possible to form a noun with *-ness*. Although occasionally the two

words have distinct meanings (e.g., *monstrosity* and *monstrousness* mean different things), much of the time the *-ity* and *-ness* words do appear to be synonymous, contrary to our hypothesis. It is therefore not possible to say that blocking always occurs to avoid synonymy

Further examples of synonymous derivation in English can be found with other sets of rival affixes. For example, for the suffixes that derive nouns from verbs that we mentioned above, COCA gives us examples like *omitment* alongside the more frequent *omission*, or *disrupture* alongside *disruption*, with no apparent difference in meaning. The affixes *-ish*, *-esque*, and *-like* are another set of rival affixes. Again, some searching in COCA turns up cases of the same base with each of these suffixes where the derived words in context seem to have the same meaning, for example, *Barbie-ish*, *Barbieesque*, and *Barbie-like* or *iPod-ish*, *iPod-esque*, and *iPodlike*.⁵ A final rival set of affixes is *-hood*, *-ship*, and *-dom* which form abstract nouns from other nouns. These can be found in COCA on bases like *guru* or *student*, with the resulting derived words *guruhood*, *guruship*, and *gurudom*, or *studenthood*, *studentship*, and *studentdom* apparently all used with the same meaning. Bauer, Lieber, and Plag (2013) give more examples like those above. The existence of examples like these calls the principle of blocking into question, at least with respect to derivational morphology.

There is also some evidence that blocking isn't a hard and fast rule in inflection either. For example, although adult English speakers always use the irregular plural *mice* when referring to furry little vermin, they vary between *mice* and *mouses* when referring to computer pointing devices (interested students can check this out for themselves in COCA). And there's a great deal of variation between the past participle forms of the verb *dive* (either *dived* or *dove*) and *forget* (*forgot*, *forgotten*). So blocking does not seem to be inevitable in inflectional forms either.

Does blocking exist? On the basis of our intuitions, morphologists have long had the feeling that it does, but looking at actual examples in a large corpus calls that feeling into question. Why do our intuitions about blocking and synonymy avoidance feel so real, then? It may be that our mental lexicons allow us easiest and quickest access to the words we are exposed to the most frequently, so that alternative forms sound odd to us. Only a combination of further study of the organization of the mental lexicon and further analysis of data from corpora are likely to shed new light on the subject.

10.5 Constraints on Affix Ordering

At various points in this book we have talked about how affixes are ordered with respect to each other. For example, in Chapter 6 we noted that inflectional affixes generally come outside of derivational affixes

5. In COCA, examples with these suffixes are sometimes hyphenated and sometimes not. Here I have preserved the hyphenation found in COCA.

in the languages of the world. We also mentioned in [Chapter 7](#) (Bybee 1985) that among inflectional affixes, tense and aspect inflection tends to come closer to the stem than person and number inflection, the reason being that tense and aspect are more closely relevant to the meaning of the verb than person and number.

In this section, we will look more closely at the issue of how English derivational affixes are ordered with respect to each other, as this has been a matter of theoretical dispute for some time. The problem is this: if the only thing that constrained the ordering of affixes in English were the categorial restrictions on their attachment (e.g., that *-ness* only attaches to adjectives, or that *-ize* attaches to both nouns and adjectives), we would expect to find many more combinations of affixes than we do find. In fact, we find very few of the potential combinations of affixes. The theoretical issue, then, is what restricts the combination of affixes, and how we explain those restrictions.

We have already touched upon this problem. In [Chapter 9](#) we noted that native derivational suffixes in English tend to come outside of non-native ones. One explanation that has been proposed for this generalization is that the two strata of English derivational morphology are ordered with respect to each other such that non-native affixation precedes native affixation. If the two strata are strictly ordered with respect to each other, then non-native affixes will never be able to attach outside of native ones. We might call this the **Stratal Ordering Hypothesis**.⁶

There are two problems with this hypothesis, however. One is that it's not quite accurate. There are non-native affixes that do not cause stress or phonological changes, like other non-native affixes, and that are perfectly happy attaching to native bases, for example *-ee* (*standee*), *-ize* (*winterize*), *-able* (*singable*). Occasionally it is even possible to attach a non-native affix to a word formed with a native suffix; the words *softenable* or *whitenable*, for example, have native *-en* followed by non-native *-able*. This has led theorists to lump *-ee*, *-ize*, and *-able* in with native affixes, thus blurring the lines between the strata.

More seriously, if the only thing which constrained the ordering of derivational affixes in English were the ordering of the two strata, we would still expect to find many more combinations of affixes than we do. For example, the suffixes *-age* and *-ize* are both non-native. The suffix *-age* forms nouns, and *-ize* attaches to nouns, so *-ize* should attach to *-age* words. But we never find words with the combination *-ageize*, and words we might coin on the spot sound quite odd (*orphanageize?*, *baggageize?*). Similarly, *-ify* forms verbs, and non-native *-ance* forms nouns from verbs, but we never get nouns like *purifiance*. So another problem for the theory of stratal ordering is that the combinations of affixes within strata are more limited than the Stratal Ordering Hypothesis would lead us to expect.

How else might we constrain the ordering of affixes? One possibility that has been proposed (Plag 1999; Giegerich 1999) is that affixes not

6. In the literature on morphology, the term Level Ordering Hypothesis has frequently been used, for reasons we do not need to go into here.

only select what they attach to (native or non-native bases of particular categories), but also what attaches to them. For example, according to this hypothesis, the reason that we don't find words like *purifiance* is that the suffix *-ify* selects the suffix *-ation* as its nominalizer. So we find *purification* (with an extender *-c-*), and we predict that any new verb in *-ify* that we create (say, *Trumpify*), will allow *-ation* to attach to form its nominalization (therefore *Trumpification*). Similarly, any verb formed with the prefix *en-* in English (e.g., *entomb*), will form its nominalization with *-ment* because *en-* selects *-ment* as its nominalizer (so *entombment*, rather than *entombal* or *entombation*). This sort of selection is called **base-driven selection**.⁷

Another proposal is called **Complexity Based Ordering**. According to Hay and Plag (2004: 571), the gist of this proposal is that “the less phonologically segmentable, the less transparent, the less frequent, and the less productive an affix is, the more resistant it will be to attaching to already affixed words.” For example, the suffix *-ness* is extremely productive, its meaning is always transparent, and it's easily segmentable from its bases. According to this hypothesis, it should not be resistant at all to attaching to other affixes, and this is indeed what we find, as the examples in (21a) show. By contrast, the verb-forming suffix *-en* (as in *shorten*, *deepen*), is not terribly productive or frequent, and because it is vowel-initial, is less easily segmentable from its base than is *-ness*, which is consonant-initial. As the hypothetical examples in (21b) show, it is difficult to find any suffixes that *-en* can attach to:

- (21) a. courtliness, amateurishness, aimlessness, carefulness
 b. *hopefulen, *happinessen, *shoelessen

Indeed the only affix that *-en* can attach to is *-th* (*lengthen*, *strengthen*), which is completely unproductive in English, and even less segmentable from its bases than *-en* itself is.

What we can see is that there are a number of hypotheses that partially explain how derivational affixes are ordered with respect to each other in English, but that this question is by no means settled. Theorists will continue to work on this problem for some time to come.

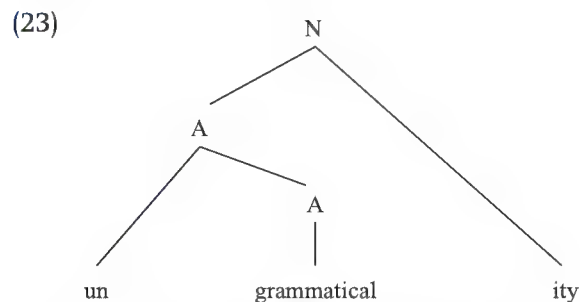
10.6 Bracketing Paradoxes

Bracketing paradoxes are cases in which either the semantic interpretation or the phonological organization of a word seems to conflict with its internal structure. Consider the words in (22):

- (22) ungrammaticality
 blue-eyed
 unhappier

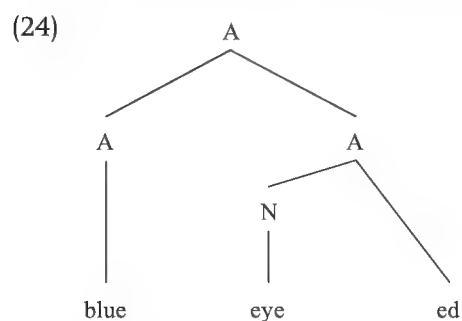
7. Note that the term *base* in this context is used in a broader sense than I have used it in this book. For Plag and Giegerich the base of a word is whatever simple or complex form an affix attaches to.

At first glance, these are unremarkable words. But if we look more closely at them, you'll see that they raise some problems. The word *ungrammaticality* needs to have the structure in (23):



Since the prefix *un-* attaches to adjectives and does not change category, it must attach first to the base *grammatical*.⁸ The suffix *-ity* attaches to adjectives and forms nouns. So the structure in (23) would be justified according to the structural requirements of the affixes. But *un-* is a native prefix, and *-ity* a non-native suffix. The structure in (23) therefore requires that a native prefix go inside a non-native suffix, contrary to the generalization we saw in Section 10.5. In other words, if there really is some sort of constraint against non-native affixes appearing outside native affixes, the word *ungrammaticality* is paradoxical.

Of course, this example may not be such a problem after all, since – as we saw above – there are other cases in which a non-native affix appears outside of a native one. The word *blue-eyed*, however, is paradoxical in a way that cannot be attributed to stratal ordering. It appears to be a compound of the adjectives *blue* and *eyed*, the second of which is itself a complex word consisting of the noun *eye* and an adjective-forming suffix *-ed*:



But the structure in (24) implies that there is an independent adjective *eyed* and that seems not to be the case. Besides, the word *blue-eyed* seems to mean 'having blue eyes', which would suggest the structure in (25), rather than the one in (24):

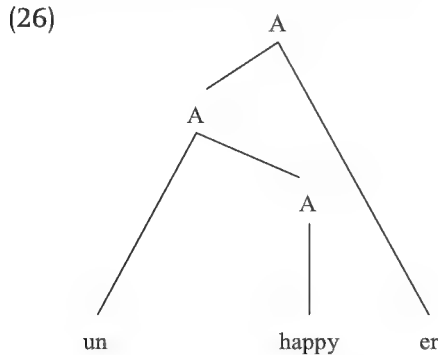
(25) [[blue eye] ed]

How do we explain this paradox? In fact, a number of solutions to this paradox have been suggested. One plausible reason is that the structure

8. This is of course itself a complex word, but as its internal structure is not relevant to the issue we're discussing here, we will ignore that internal structure.

in (24) is in fact correct, and that there is a pragmatic reason why we don't find an independent word *eyed*. We don't find such a word because it's usually not a useful concept. People assume that living organisms have eyes, so we'd never have a reason to point out the 'eyed one' as opposed to the one without eyes. But given a context in which such a contrast is plausible – say, comparing two space aliens – it no longer seems so absurd to think of an independent word *eyed*.

Our final example of a bracketing paradox is the word *unhappier*, which is paradoxical for yet a different reason. Here, the semantic interpretation of the word would suggest the structure in (26):



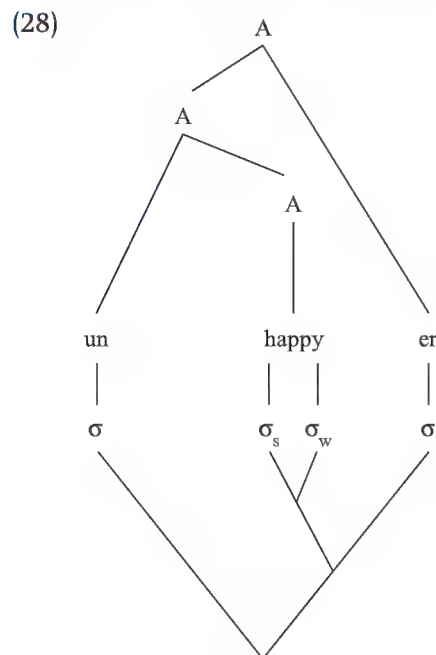
The word *unhappier* seems to mean 'more unhappy', with the comparative suffix *-er* applying to the negated adjective, so the semantic interpretation of the word corresponds to this structure. However, as we saw in [Chapter 8](#), the comparative suffix *-er* has a phonological restriction on its attachment that calls the structure in (26) into question. Recall that the comparative suffix attaches to one-syllable adjectives, and to two-syllable adjectives whose second syllable is unstressed. Two-syllable adjectives whose stress falls on the second syllable (e.g., *aghost* or *upset*) cannot take the comparative *-er* suffix, but generally prefer the periphrastic comparative (*more aghost*, *more upset*). And three-syllable adjectives almost never form their comparatives with *-er*. The problem with the structure in (26), then, is that *-er* looks like it has attached to the complex adjective *unhappy*, which consists of three syllables. Of course, if the *-er* were first attached to *happy* and then *un-* attached outside that, as in (27), there would be no problem:

(27) [un [[happy]er]]

But this does not accurately reflect the meaning of the word. The structure in (27) suggests that the word means 'not happier' rather than 'more unhappy'.

Here too, there is a potential solution to the paradox. Many morphologists believe that words have two separate structures, one which reflects the syntax and semantics of the word, and a separate structure which reflects the prosodic organization of the word into syllables, feet, and higher levels of phonological organization. We will not go into the details of different levels of phonological organization here, but I can

give you just a suggestion of what I mean. Suppose that the word *unhappier* has two simultaneous structures:



The top structure is identical to the one in (27) and reflects the semantic organization of the word. The one on the bottom reflects the phonological organization of the word, where σ stands for 'syllable', σ_s for 'stressed syllable', and σ_w for 'unstressed syllable'. If words are allowed to have two separate and simultaneous representations, one for their syntactic and semantic structure, and another for their phonological structure, the word *unhappier* is no longer paradoxical.

10.7 The Nature of Affixal Polysemy

Unlike the problems of affix order and bracketing paradoxes, which have received a great deal of attention from morphologists, the problem we will take on in this section has received little attention, but it is no less interesting. In [Chapter 3](#) we briefly touched upon the issue of affixal polysemy. To refresh your memory, affixal polysemy is the tendency for affixes to have several closely related meanings. For example, we pointed out in [Chapter 3](#) that it is a curious fact that the affix that is used for making agent nouns in languages is frequently also used for making instrument nouns. As I pointed out, this is the case in English, and Dutch – not surprising, as they are closely related languages – but also in Yoruba (Niger-Congo), Turkish (Turkic) ([Lewis 1967: 225](#)), Kannada (Dravidian) ([Sridhar 1990: 273](#)), and many other languages. It cannot be an accident that agent nouns and instrument nouns are so often created by the same affixes. So the theoretical question that arises is why this should be. To answer this question, let's again look a bit more closely at English.

(29)	<i>-er</i>	agent	writer, driver, thinker, walker
		instrument	opener, printer, pager
		experiencer	hearer
		patient/theme	fryer, sinker
	<i>-ant/-ent</i>	agent	accountant, claimant, servant
		instrument	adulterant, irritant
		experiencer	discernant
		patient/theme	descendant

It appears that in English the suffix *-er* forms not only agent and instrument nouns, but also nouns that denote the experiencer of an action, or even the patient or theme of an action. Even more curious, the suffix *-ant* covers the same range of meanings. The theoretical question we must raise in light of these data is why *-er* and *-ant* nouns cover just this range of meanings and not some others.

The first step in explaining this affixal polysemy is to figure out just what *-er* and *-ant* mean. One suggestion that has been made (Lieber 2004) is that these affixes don't actually mean 'agent' or 'instrument', but something much more abstract – something like 'concrete noun concerned with a process or event'. In this, they are semantically analogous to simple nouns like *poet* or *awl* that denote people or things defined by what they do. One piece of evidence for this claim is that *-er* can attach to nouns as well as to verbs, and when it does, it always adds an active or eventive element of meaning to its noun base. So, for example, a *villager* is someone who lives in a village, and a *freighter* is something that carries freight. No verb is necessary for the active part of the meaning – this comes directly from the suffix.

The second part of the answer to our question has to do with understanding the argument structures of verbs. You'll recall from Chapter 8 that the argument structure of a verb consists of those arguments that are semantically necessary to the verb (see Section 8.2). What is most interesting for our purposes is that in English there is a range of semantic roles that the subject of a verb can play:

- (30)
- | | | |
|----|----------------------------------|---------------|
| a. | Fenster ate the pizza. | (agent) |
| b. | The key opened the door. | (instrument) |
| c. | We heard the neighbors fighting. | (experiencer) |
| d. | The boat sank rapidly. | (theme) |

In (30a), the subject of the sentence is the agent or 'doer' of the action. In (30b), the subject is called an **instrument** rather than an agent, because it is an inanimate noun that does something. In (30c), we call the semantic role that the subject plays the **experiencer** rather than agent, because one can hear something without doing anything at all – to be an agent, one must act intentionally. Finally, the subject in (30d) is not the agent, but the theme, in other words, the noun that undergoes or is moved by the action.

What is interesting is that *-er* and *-ant* nouns denote exactly the semantic role conveyed by the subject of their base verb. So an *eater* is an agent,

just as the subject of the verb *eat* is an agent, an *opener* is an instrument, just as the subject of *open* is an instrument, a *hearer* is an experiencer, just as the subject of *hear* is an experiencer, and a *sinker* is a theme, just as the subject of *sink* (intransitive) is a theme. The reason that *-er* and *-ant* display exactly the range of meanings that they do is that they are linked to the subject argument of the verb. They can mean whatever the subject of a verb can mean.

10.8 Reprise: What's Theory?

The few topics I've touched upon here just barely scratch the surface of theoretical questions that linguists can raise about morphology. Most of the issues we've looked at in this chapter concern English, although if we had time to delve further into them, we'd see that they concern many other languages as well. What is important to keep in mind, though, is that there are many theoretical issues that come up only when we look beyond English to the analysis of other languages.

Summary

In this chapter we have first considered what we mean by morphological theory, and then explored a number of theoretical topics that have been important to generative morphologists over the years. We have looked at the nature of morphological rules and the predictions specific models of morphology make about the sorts of morphology we ought to find in the languages of the world. We have explored the issue of the relation between morphology and syntax and the extent to which these two levels of linguistic organization can be kept separate from one another. We have looked at how the patterns of ordering in derivational affixes can be explained, how bracketing paradoxes can be resolved and how affixal polysemy can be explained.

Exercises

1. Consider the sort of templatic morphology that we looked at in [Section 5.6](#). Do you think that templatic morphology presents any problems for Item and Arrangement theories of morphology?
2. Consider the verb *shit*. What are the past tense and past participle forms of this verb. Check a dictionary if you're not sure. Does this verb give us any evidence for or against blocking?
3. Consider the words *pomp*, *pompous*, and *pomposity*. Do they offer evidence for or against blocking?
4. Why might the following words be considered bracketing paradoxes?
three-wheeler
whitewashed

transformational grammarian
nuclear physicist

5. In [Section 10.7](#) I suggested that the suffix *-er* can have whatever semantic role is carried by the subject of its base verb. Consider the words *loaner* and *keeper* in the sentences below:

- i. My car was in the garage so they gave me a loaner.
- ii. This book is a keeper.

What challenge do these forms present for the hypothesis in 10.7?

6. The suffix *-ee* is usually said to form 'patient nouns', that is, nouns that denote the person who undergoes or is subject to the action denoted by the base verb. Consider the following examples, and discuss the extent to which *-ee* exhibits affixal polysemy:

employee
nominee
standee
escapee
addressee
amputee

7. Consider the following prefixed words (from [Bauer, Lieber, and Plag 2013: 600](#), data from COCA). Each word has two prefixes. First use the online *OED* or another dictionary to identify which prefixes are native and which are non-native. Then discuss what problems, if any, these examples pose for the Stratal Ordering Hypothesis.

ex-betrothed
hyper-unemployment
mini-outbreak
polyunsaturated
post-midnight

11 Theories of Morphology

CHAPTER OUTLINE

KEY TERMS

Vocabulary Item

Spell Out

Late Insertion

construction

schema

Paradigm

Function

Realization Rule

naturalness

learning

mechanism

cue

outcome

lexical semantics

semantic

skeleton

semantic body

In this chapter you will read about six different current theories of morphology, looking at their philosophical bases, their main concerns or leading questions, and at where they stand on a couple of the prominent questions with which morphologists have been concerned, specifically, the balance between storage *versus* rules and the status of the morpheme. The theories we will look at are:

- ◆ Distributed Morphology
- ◆ Construction Morphology
- ◆ Paradigm Function Morphology
- ◆ Natural Morphology
- ◆ Naïve Discriminative Learning
- ◆ The Lexical Semantic Framework

After reading this chapter you should be ready to start reading primary literature in morphological theory for yourself.

11.1 Introduction

In this text, you have learned about the data that morphology is concerned with, about the distinctions morphologists make between derivation and inflection, about morphological typology, about the relationship between morphology on the one hand and syntax and phonology on the other, among other things. I started from the assumption that studying morphology is a way of understanding and modeling the mental lexicon, answering such questions as what the contents of the mental lexicon are and how those contents are organized. In [Chapter 10](#) we looked at a number of basic theoretical issues that have fascinated morphologists over the last few decades.

The study of morphology involves more than analyzing data, though, and morphological theory is concerned with more than just a handful of disparate theoretical issues. At its best, morphological theory is a way of conceptualizing and explaining our lexical knowledge and our ability to build words in our language. Theories of morphology all seek the nature of our ability to build new words, but they often start with different leading questions and focus on different aspects of lexical knowledge. Depending on how they answer some basic philosophical questions and on what sort of data a theorist is most interested in, theories or frameworks of morphology may end up giving rather different pictures of lexical knowledge. In this chapter we will explore a number of contemporary theories of morphology, looking at the main questions that they ask, the assumptions that they make about the nature of language in general, and the sorts of empirical data that they seek to explain.

I think it's safe to say that most theories of morphology today are concerned in some way with the nature of the mental lexicon. We can begin to understand their differences by asking several key questions:

- Does this theory take our knowledge of language to be distinct from that of other cognitive systems?
- Is this theory concerned primarily with the formal aspects of morphological knowledge (what the rules of inflection and word formation look like) or with its function (how complex words work, why they are the way they are)?
- Does this theory take the mental lexicon and our means of adding to it to be distinct from our syntactic and phonological knowledge?
- What is the central question or type of data driving the architecture of this theory? Is it driven by a specific empirical question (say, the nature of inflectional paradigms or affixal polysemy), by a formal goal (simplicity, elegance), by a psycholinguistic question (how is morphology learned or processed)?

In the next six sections, I will offer a brief survey of some of the theories that have been prominent in the early part of the twenty-first century. As we go through them, there are several things to keep in

mind. First, these theories (indeed all theories!) are works in progress. There are multiple theorists working in each framework. As theorists examine new data and refine a theory, it develops and changes – theories are always moving targets. You should also keep in mind that my synopses are of necessity brief and cannot hope to give you more than a flavor of what the theory seeks to do. In choosing what to highlight about each theory I will inevitably miss out on something that theorists working within that framework think is important – I hope that those theorists will not feel misrepresented or short-changed. You should keep in mind as well that the frameworks I highlight here are not the only ones out there, nor are they necessarily the best ones. I have chosen them to illustrate the different ways in which the questions that we ask about words and word-building lead us to very different views of the mental lexicon. Finally, in what follows I will not judge theories against one another. Linguists can be quite partisan in arguing for or against specific theories (myself included!), but my goal here is to familiarize you with a range of theoretical models rather than to advocate for one or another.

A caveat before we go on. As we saw in [Chapters 8 and 9](#), it is difficult to separate the field of morphology from either matters of syntax or of phonology. Because of the extent of this interconnectedness, I will sometimes need to refer to aspects of syntactic and phonological theory that you may not yet have encountered in your studies; you may therefore find it necessary to consult a current textbook in syntax or phonology to fill in material you might not have yet been exposed to.

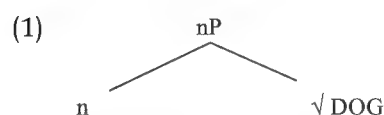
11.2 Distributed Morphology

Distributed Morphology – referred to just by the initials DM – is a theory of morphology that has arisen out of the general framework of generative grammar, specifically out of Chomsky's Principles and Parameters and subsequent Minimalist theory of syntax ([Chomsky 1981, 1995](#)). DM shares with generative grammar the philosophical commitment to a mentalist treatment of language: the goal of generative grammar is to model the language faculty, that is, the rules that allow speakers of a language to produce and understand infinite numbers of sentences, and with respect to morphology, words. And as with generative grammar, DM is committed to the belief that linguistic knowledge is specialized and distinct in form from other areas of cognition.

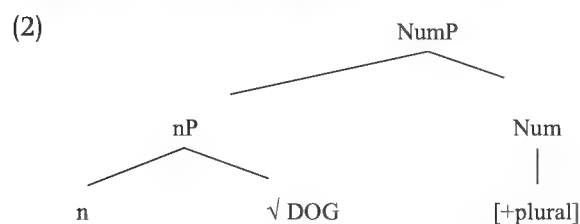
As with generative grammar in general, DM is a formal theory, concerned with the organization and architecture of the rules that generate complex words ([Halle and Marantz 1993](#), [Harley and Noyer 1999](#), [McGinnis-Archibald 2016](#), [Siddiqi 2019](#)). And as with generative grammar as a whole, it is a theory that values simplicity and elegance ([Siddiqi 2019](#): 152). Discussing his theory of syntax, [Chomsky \(1993: 2\)](#) says: “A working hypothesis in generative grammar has been that languages are based on simple principles that interact to form often intricate

structures, and that the language faculty is nonredundant.” For DM, this commitment to simplicity leads to the assumption that there is a single “generative engine” in the grammar. What this means is that the theory contains only one way of building structure and that way must generate both sentences and complex words. To put it succinctly, in DM morphology simply is syntax. The rules that build sentences also build derived and inflected words.

In Minimalist Syntax and therefore in DM, there are two rules that build structure, Merge and Move. We will concentrate on Merge here. In DM Merge builds binary-branching structures that combine two sorts of elements. These elements are not words or morphemes per se, but rather are either morphosyntactic features like [+/-plural] or [+/-past] or Roots (Harley 2009).¹ Roots are abstract in the sense that they do not have either syntactic category or phonological content.² For example, the root $\sqrt{\text{DOG}}$ is neither a noun nor a verb, but could turn out to be either (*The dogs barked*, *The problem dogged me*) depending on the syntactic structure in which it finds itself. In order for the root $\sqrt{\text{DOG}}$ to become a noun, it needs to be categorized by a node called ‘little n’.



If this structure is then merged with the number feature [+plural], we would get a structure like (2), where NumP stands for ‘Number Phrase’:



This structure, in turn, might be merged with features like [+definite] as part of a determiner phrase (DP) and so on. DM is in some ways like an Item and Arrangement system in that the terminal nodes of syntactic trees are ‘pieces’ that are somewhat like morphemes. But they are not classical morphemes because they have no phonological form.

It is only after all syntactic structure is built that the terminal nodes of trees like those in (1) and (2) are associated with phonological material.³ The phonological bits that get associated with the terminal nodes of

1. Roots are sometimes called l-morphemes (Harley and Noyer 1999). Syntactic nodes can sometimes contain bundles of features, that is, several features at once.

2. Indeed, in some versions of DM, they don’t have semantic content either (Siddiqi 2019: 163).

3. In some versions of DM there is a stage in the derivation after all syntactic operations have been completed but before Vocabulary Insertion. At this stage, called Morphological Structure, morphological features can be added (for example, those that have no syntactic effect), deleted, or fused together, or nodes in a tree can be reordered (Siddiqi 2019: 148).

trees are called **Vocabulary Items**, and the process of associating them is called **Spell Out**. Because this step happens only after all syntactic operations have finished, it is sometimes referred to as a **Late Insertion** model. In the case of the tree in (2), the root $\sqrt{\text{DOG}}$ would be associated with the phonological string /dɔg/ and the feature [+plural] with the string /z/. Each Vocabulary Item is “a relation between a phonological string or ‘piece’ and information about where that piece may be inserted” (Harley and Noyer 1999: 4). For the plural /z/ this relation might be represented as in (3):

(3) /z/ ↔ [+plural]

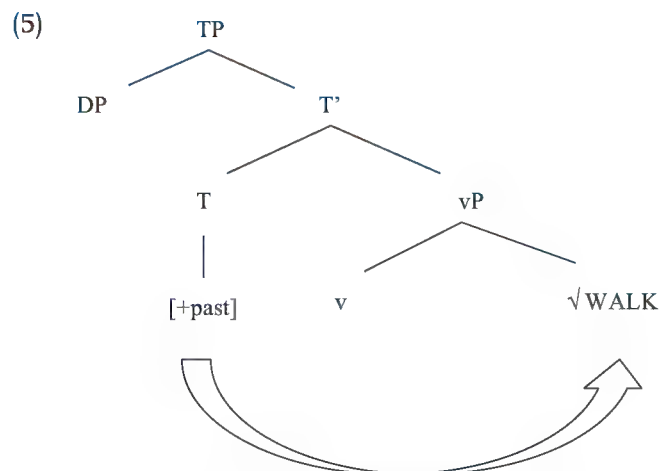
Vocabulary Items are also associated with meanings, which are to be found in what DM theorists call the **Encyclopedia**. The leading idea behind DM is that the traditional notion of a lexical item – a pairing of sound and meaning – is distributed in several places in the grammatical model: the Roots and morphosyntactic features in the syntactic tree, the Vocabulary Items in a list that we tap into after the syntactic derivation, and an Encyclopedia that associates Vocabulary Items with their conceptual content.

With a regular plural noun like *dogs*, there is a one-to-one correspondence between the terminal nodes in the syntactic tree and the Vocabulary Items that get inserted. There are a number of ways in which this simple scenario can be complicated, however, and these illustrate some of the important features of DM. Suppose, for example, instead of $\sqrt{\text{DOG}}$, we had the root $\sqrt{\text{OX}}$. In syntactic structure, the representation of the plural of $\sqrt{\text{OX}}$ would look the same as the plural of $\sqrt{\text{DOG}}$, which captures the intuition that all plural nouns behave alike with respect to syntax, regardless of how the plural is realized phonologically. However, at the point of Spell Out, we would not want the Vocabulary Insertion rule in (3) to associate /z/ with $\sqrt{\text{OX}}$ + [+plural]. Instead, as the correspondent to [+plural] for $\sqrt{\text{OX}}$ is /ən/, we would need a special Vocabulary Insertion rule, maybe something like (4) (modeled after Vocabulary Insertion rules in Halle and Marantz 1993):

(4) /ən/ ↔ X + [+plural], where X = /aks/

An important principle of DM is that a specific rule takes precedence over a less specific rule. Since (4) contains more specific information than (3) does, making reference to a specific Vocabulary Item /aks/, /ən/ would be inserted before /z/ would have a chance to be. Another important principle of DM is that Vocabulary Insertion rules are realizational in nature. For example, if the terminal strings in our tree had the root $\sqrt{\text{MOUSE}}$ + [+plural], we might have a rule that fused these nodes together and replaced them by the single phonological string /maɪs/.

DM also allows the use of movement rules to rearrange syntactic structure. For example, in an English sentence with only a main verb a well-known syntactic rule, sometimes called T to V, lowers a tense feature onto a verb root $\sqrt{\text{WALK}}$, as in (5). At the end of the syntactic derivation, the complex of the root plus the feature [+past] can then undergo Spell Out:



There is much more to be said about how DM analyzes various kinds of data and we cannot hope to do any more than give you the flavor of DM analyses here. But it is worth pointing out two important characteristics of DM. First, DM does not make a strict distinction between inflection and derivation; what are traditionally called inflectional and derivational morphemes are simply bundles of features in DM (Siddiqi 2019: 161). Interestingly, however, DM so far has had much more to say about inflection than about derivation or compounding. As we will see in [Section 11.6](#), other frameworks also concentrate their energy on matters of inflection, and make important theoretical use of the notion of ‘paradigm’. DM, however, does not make use of the paradigm as a theoretical construct – rather, inflectional forms emerge from syntactic structure, but follow no principles of organization independent of that syntactic structure. Paradigms don’t really exist in DM.

A second point is that although DM is a mentalist theory, it does not encompass a ‘mental lexicon’ per se. Being part of the Chomskian generative tradition, DM is a theory of mental representation, that is, the rules that speakers know implicitly that allow them to produce complex words and sentences. DM theorists don’t believe that complex words are stored; everything is generated on line as part of the syntactic derivation. Our mental representations of complex words thus consist of rules like Merge and Move, a list of Vocabulary Items, an Encyclopedia, and various morphological rules that manipulate the roots and features of syntactic trees to prepare them for Vocabulary Insertion. There is no mental lexicon per se, if what we mean by ‘mental lexicon’ is a repository of words and rules for forming words. For DM, our knowledge of words – simple and complex – is simply an artifact of several dispersed – distributed – components of our grammatical knowledge.

11.3 Construction Morphology

Construction Morphology, sometimes abbreviated as CxM, is a branch of a more general theory called Construction Grammar, abbreviated CxG (Booij 2010, 2016; Audring and Masini 2019). CxG and CxM are, like DM,

formal theories, but their assumptions about form are rather different from that of DM. Construction Grammar (Goldberg 1995) holds that our mental representations of language are composed of pairings between form and meaning. Individual simple words, of course, are easily conceived of as pairings between a form and a meaning – what is traditionally referred to as Saussurean signs (Saussure 1972/1983), but in CxG, these pairings can extend to whole syntactic constructions. A **grammatical construction** is a pattern that imparts meaning apart from the words that make it up. So for example, the verb *whistle* as we normally use it denotes a certain sort of sound emission (*Fenster whistled a catchy tune*). However when inserted in what is called the ‘X’s Way’ construction, it comes to mean a manner of moving:

- (6) *X’s Way Construction*: X verb-ed X’s way PP.
Fenster whistled his way out of the kitchen.

The movement meaning arises from the construction as a whole, rather than from the verb *whistle*.

CxM extends the notion of construction to morphology, hypothesizing that the mental lexicon is comprised of a complex network of constructions which might be simple words like *giraffe*, complex words like *happiness*, multiword units like idioms (*kick the bucket*), or syntactic constructions like *X’s Way*.

Looking specifically at morphology, constructions represent generalizations over forms that are part of our lexicon, but they are also responsible for generating new forms. Let’s look at a specific example. In learning English a speaker might pick up simple words like *happy*, *red*, and *ugly*, and develop a mental representation for them that looks something like (7):

- (7) /hæpi/ ↔ [happy]_A ↔ HAPPY
/red/ ↔ [red]_A ↔ RED
/ʌgli/ ↔ [ugly]_A ↔ UGLY

In the representations in (7), these words are shown as the linking of a phonological form /hæpi/, a morphosyntactic form [happy]_A, and a semantic representation HAPPY (the latter, is represented informally in capital letters in this theory). The double arrow shows the linking between one part of the representation and another. If a speaker also hears words such as *happiness*, *redness*, and *ugliness*, they may create mental representations for those forms as well:

- (8) /hæpinəs/ ↔ [[happy]_A ness]_N ↔ THE STATE OF BEING HAPPY
/rednəs/ ↔ [[red]_A ness]_N ↔ THE STATE OF BEING RED
/ʌɡlɪnəs/ ↔ [[ugly]_A ness]_N ↔ THE STATE OF BEING UGLY

If the speaker is exposed to enough examples, a pattern begins to emerge that receives its own entry in the mental lexicon, perhaps something like (9):

- (9) [[X]_A ness]_N ↔ THE STATE OF BEING X_A

The representation in (9) is what is called a **schema** in CxM. Schemas emerge as generalizations over forms in our mental lexicons, but once they've emerged, they can then also serve to generate new forms. So for example, if a speaker hears a new adjective like *yummy*, they can use the schema in (9) to create *yumminess*.

CxG conceives of the mental lexicon as a complex and rich network of connections among words and schemas. Some schemas are arranged hierarchically. As speakers build their mental lexicon, they may notice that there are words that give rise to schemas that are very similar to (9), for example, (10), which might emerge from learning forms like *purity*, *curiosity*, and *obesity*.

$$(10) \quad [[X]_A \text{ity}]_N \leftrightarrow \text{THE STATE OF BEING } X_A$$

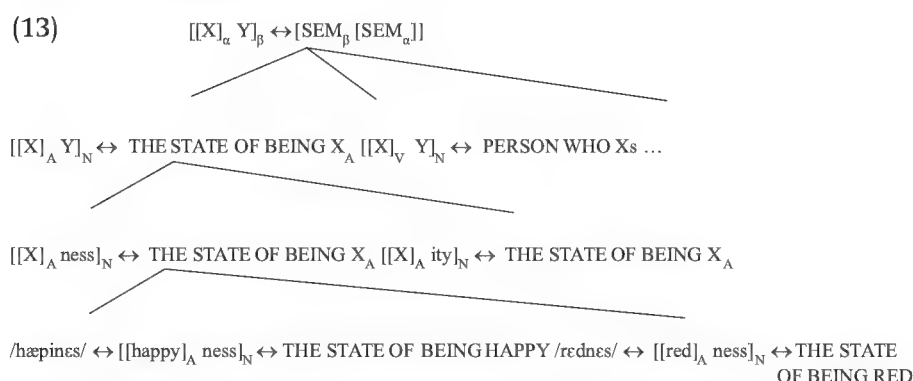
The difference between (9) and (10) lies in the phonological form that follows the base. Generalizing over the schemas in (9) and (10), the speaker may then begin to develop the schema in (11):

$$(11) \quad [[X]_A Y]_N \leftrightarrow \text{THE STATE OF BEING } X_A$$

And if a speaker then creates schemas for suffixes like *-er*, *-ation*, *-hood*, and so on, an even more general schema for suffixation in English might begin to emerge (Masini and Audring 2019: 370).⁴

$$(12) \quad [[X]_\alpha Y]_\beta \leftrightarrow [\text{SEM}_\beta [\text{SEM}_\alpha]]$$

These schemas can be arranged in a hierarchy, with the most general at the top, and increasingly specific ones below:



Schemas can be linked in other ways as well. For example, suppose that English has the schemas in (14) for the prefix *de-* and the suffix *-ize*:

$$(14) \quad [\text{de}[X]_{N,V}]_V \leftrightarrow \text{REMOVE } X \text{ FROM, UNDO } X_{N,V}$$

$$[[X]_{A,N} \text{ize}]_V \leftrightarrow \text{MAKE, CAUSE TO BE } X_{A,N}$$

These schemas would allow us to form verbs like *deice* and *debug* or *unionize* and *equalize*. Suppose, however, a speaker encounters the word

4. SEM stands for the semantic representations of X and Y, whatever they are, and α and β stand for syntactic categories like N, A, and V.

denuclearize. In a theory that has traditional word formation rules, we would first have to create the verb *nuclearize* and then attach the prefix *de-* to get *denuclearize*. Classic word formation rules are linear and procedural: they apply one after the other like an assembly line. But our speaker may never have heard the word *nuclearize*. Indeed, although *nuclearize* is a possible word of English, it's one that rarely, if ever, occurs. It seems odd for our mental grammar to create a word that does not exist for us to get to a word that does. For this reason, CxM does not represent complex words in terms of ordered word formation rules. It does not force us to postulate the creation of a non-existing word to get to an existing word. Rather, it allows us to put together or 'unify' the two schemas in (14) to give the schema in (15):

$$(15) \quad [de[X]] + [[X]ize] = [de[X]ize]$$

The mental representation for *denuclearize* can then be (16), without there being an intermediate step at which a verb *nuclearize* has to occur.

$$(16) \quad [de[nuclear-ize]] \leftrightarrow [UNDO [MAKE NUCLEAR]]$$

So far, our examples have come from word formation, but CxM handles inflection in the same way it handles derivation and compounding, that is, using schemas. The schema for plural nouns in English, for example, might look like (17), where x_i is the phonological representation of N_i and SEM_i is the semantic representation of the noun:

$$(17) \quad /x_i -z/ \leftrightarrow [N_i, +pl] \leftrightarrow [PL [SEM_i]]$$

CxM can also express paradigmatic relationships in either derivation or inflection using what are called **second order schemas**. In second order schemas, two or more schemas are linked or related to each other in our mental lexicon, the symbol for this linking being $\approx$. (18) shows a paradigmatic linking between schemas for the singular noun *dog* and the plural noun *dogs* in English:

$$(18) \quad /dɔg_i/ \leftrightarrow [N_i, +sg] \leftrightarrow [SG [SEM_i]] \approx /dɔg_i -z/ \leftrightarrow [N_i, +pl] \leftrightarrow [PL [SEM_i]]$$

For inflectionally related forms, a series of schemas related by $\approx$ s would be the equivalent of an inflectional paradigm. For derived words, schemas related by $\approx$ s might be called a word family (or a derivational paradigm).

There are two general observations that we can make about CxM. First, this framework does not recognize morphemes as independent parts of the mental lexicon. What are classically called morphemes do not exist as entities in our mental lexicons, but rather are incorporated as parts of schemas. Second, CxM assumes that simple words, complex words, and indeed multiword units, are stored in the mental lexicon. Schemas serve as generalizations over what is stored and allow speakers to add to their store of complex forms. Second order relationships among schemas serve as a proxy for paradigms.

11.4 Paradigm Function Morphology

Paradigm Function Morphology (PFM) has as its central goal to account for the properties of complex inflectional paradigms (Stump 2001, 2016, 2019; Bonami and Stump 2016), specifically to account for the range of variation that might be possible across inflectional systems in the languages of the world. In other words, it has a sort of typological focus that is unique, compared with the other theories we have looked at so far. For proponents of PFM, the paradigm is the quintessential architectural feature of morphology that distinguishes it from syntax and phonology.

PFM starts with the assumption that lexemes are associated with sets of word forms that together comprise a paradigm.⁵ Each word form of a lexeme instantiates a cell in a paradigm that associates a root with a set of morphosyntactic properties. The rules that associate the root plus its morphosyntactic properties in that cell is called a **Paradigm Function**. For example, the past tenses of the verbs *walk* and *go* in English, would have the Paradigm Functions in (19) (Stump 2019: 287):

- (19) a. PF (<*walk*, {finite past}>) = <*walked*, {finite past}>
 b. PF (<*go*, {finite past}>) = <*went*, {finite past}>

(19a) states that the lexeme *walk* is mapped onto the word form *walked* when it bears the feature {finite past}. (19b) says that the lexeme *go* is mapped onto the word form *went* when it bears the feature {finite past}.

In PFM, **Realization Rules** are larger generalizations over Paradigm Functions. Some Realization Rules are **Rules of Exponence**, which associate roots that have specific morphosyntactic features with the phonological material that marks those features. So, the Rule of Exponence for regular past tense verbs in English might look like (20):

- (20) $X, \{\text{finite past}\} \rightarrow Xed$

Importantly, the Rule of Exponence in (20) does not presuppose that there is a past tense morpheme *-ed* that has any independent representation in the mental lexicon. It simply associates the form *X* with the form *Xed* when *X* has the features {finite past}.

The other sort of Realization Rule that PFM makes use of is called a **Rule of Referral**. Rules of Referral tell us that a word form with one set of morphosyntactic features might systematically be identical to a word form with a different set. So, for regular verb lexemes in English in which the past participle form is identical to the past tense form (like *walk*), there might be a Rule of Referral like (21) (Stump 2019: 288):

- (21) $X_v, \{\text{past participle}\} = \{\text{finite past}\}$

5. The version of PFM that I present here is what Stump (2019) calls PFM1. He refers to his more recent development of that theory as PFM2 (2016). PFM2 recognizes three sorts of interlocking paradigms: one that sets forth the content of each paradigm cell in terms of its morphosyntactic features; one that relates content cells to form cells, which may or may not be identical; and one that associates the form cells with their phonological realizations. Since the motivation for PFM2 is difficult to understand without first understanding PFM1, I present the earlier version here.

The rule in (21) says that to realize the form of a verb with the feature {past participle}, we use the rule for a verb with the feature {finite past}.

Realization rules may be arranged into blocks in which only one rule can apply to each lexeme. Of the available rules, the one that applies for a given lexeme is the most specific one. For example, suppose that there is a Rule of Exponence for past participles that only applies to a subset of verbs in English, which we will call Class *n* (Stump 2019: 288):

$$(22) \quad X_{[V, \text{Class } n]}, \{\text{past participle}\} \rightarrow Xen$$

The rule in (22) says that the past participle of a verb *X* belonging to Class *n* is realized with the form *Xen*. If the verb *eat* is a member of Class *n*, then this more specific rule will preempt the general rule in (21) which is not restricted to a particular class. The past participle form of *eat* will be realized as *eaten* rather than *eated*.

PFM is designed to account for certain important properties that appear in inflectional paradigms in the languages of the world, among them syncretism, overabundance, and defectiveness. Remember from Chapter 6 that syncretism is when one cell in a paradigm is always the same as another cell. For example, in Chapter 6 we noted that in the Slavic language Sorbian the Locative and Dative forms are always the same in the Singular and Dual (see Section 6.4.2). In Sorbian, such syncretism might be expressed by a rule that says something roughly like that in (23):

$$(23) \quad X_{[N]}, \{\text{singular, locative}\} = \{\text{singular, dative}\}$$

In other words, one way of expressing syncretism is through a Rule of Referral.⁶

We also looked briefly at defectiveness and overabundance in Chapter 6. Both occur in cases where the relationship between a cell in a paradigm and the form that fills it is not one-to-one. In the case of defectiveness we simply fail to find an inflected form that fills a cell in the lexeme's paradigm. In PFM terms, the mental lexicon would fail to have a Paradigm Function for that cell. Overabundance occurs when there is more than one form that fills a paradigm cell. In PFM, this means that the Paradigm Function for <diver, {finite past}> would have to allow two different realizations.

What about non-inflectional morphology? As a theory of inflectional paradigms, PFM has had little so far to say about derivation and compounding. Stump (2019: 299ff) notes, however, that to the extent that derived words cluster into families that seem a bit like paradigms, the theory might be extended to account for them. For example, he points out that for a verb lexeme like *inflate*, there is an intransitive form (24a), a causative form (24b), a nominalization (24c), and what Stump calls a passive potential, that is, an *-able* adjective (24d):

6. This is a bit of a simplification. See Bonami and Stump (2016: 459) for a more detailed and complete account of syncretism.

- (24) a. The mattress inflates automatically.
 b. We inflated the mattress.
 c. The inflation of the mattress takes only a minute.
 d. That mattress is inflatable.

If we consider these forms to be a small derivational paradigm, we can create paradigm functions for them.

Stump is careful to point out however, that there are a number of obstacles to overcome in assimilating derivation into PFM. For one thing, there are many verbs that would have to leave one or more of the cells of this derivational paradigm unfilled. The verb *bloom*, for example has only an intransitive form. The verb *throw* lacks an intransitive form and has no *-ation* nominalization (although there is a noun *throw* formed by conversion). In other words, whereas defectiveness and overabundance are rare features of inflection, they would be very frequent in derived words if we analyzed derived words in terms of paradigms. Further, the meanings of derived words are often too idiosyncratic to capture with a system of inflection-like features like {causative} or {passive potential}.

PFM is unlike DM and CxM in its distinctly typological focus and also in that it holds that there are principles of morphology that are distinct in form from those of syntax or phonology. PFM does, however, share some characteristics with DM and CxG. As with both DM and CxG, PFM is formal theory, concerned with the architecture of the grammar. Like DM and CxG it is mentalist, in the sense that it is concerned with the mental representations that form our grammatical knowledge. And like CxG and DM, it does not recognize the classical morpheme as a unit of analysis. Finally, PFM is unlike DM but like CxG in that it allows for both storage of lexical items and for rules that both express generalizations over those items and allow for the creation of new ones.

11.5 Natural Morphology

Natural Morphology (NM) is rather different from the theoretical frameworks that we have looked at so far in that its orientation is functional rather than formal (Dressler 1985; Bauer 2003; Dressler and Kilani-Schoch 2016; Gaeta 2019). Rather than concentrating on the architecture of our grammar or the form that rules take, NM asks the question of why languages in general and individual languages in particular are the way they are in terms of their morphology. NM posits that the answer to this question lies in general cognitive and semiotic forces that drive language to be the way it is. In particular, NM assumes that language and languages are shaped by a preference for **naturalness**, by which they mean characteristics that are most cross-linguistically frequent, most easily acquired by children, most easily processed, most resistant to language change, and least affected by disease or trauma that causes damage to the language faculty.

For NM, there are three levels on which naturalness manifests itself. First, there are universal principles of naturalness, overarching principles that are pertinent for all languages. Then there are typological preferences – certain patterns of preference that arise when some of the universal principles of naturalness are privileged over others. And finally, there are language-specific principles of naturalness, those that help to shape what is unique to individual languages. We will look at each of these levels in turn.

NM assesses morphological processes on the basis of several principles of naturalness, among them iconicity, indexicality, transparency, and biuniqueness.⁷ I will give an example of each of these in turn.

Iconicity is a relationship in which meaning and form are similar or analogous in some way. According to NM, one way in which iconicity manifests itself in morphology is the cross-linguistic preference for affixation over other morphological operations like conversion or subtraction. What is iconic about affixation is that an addition in meaning is accompanied by an addition in form. Very simply, with affixation more meaning is accompanied by more phonological (or graphemic) material. Conversion, in contrast, adds meaning without adding form and subtractive morphology adds meaning by taking away form – both of which are non-iconic and therefore less preferred in the languages of the world.

Indexicality is a universal preference for having a morpheme that modifies a base be as close as possible to that base. The more relevant a morpheme is to the meaning of a base, the closer it should be. One consequence we can derive from this preference is the strong tendency for derivational morphology to be positioned closer to the base of the word than inflectional morphology. Since inflection is more relevant to syntactic relationships outside of the immediate word, it can be closer to the outer edge of the word than derivation. Since derivation changes the part of speech of the base or adds non-grammatical meaning, it needs to be closer to the base. According to Dressler and Kilani Schoch (2016: 366), indexicality also favors fusional inflection over agglutinative. In order to get added meanings as close to the base as possible, several inflectional exponents or features might need to be bundled together, thus giving rise to inflectional fusion.

NM also assumes a universal preference for **transparency**. Morphological processes are transparent to the extent that the meanings and phonological forms of complex words are compositional and easily separable. Compositionality of meaning is sometimes referred to as morphosyntactic transparency, and separability of the phonological side of morphemes as morphotactic transparency. In English, for example, the suffix *-ness* is both morphosyntactically and morphotactically quite transparent, its non-native counterpart *-ity* rather less so. That is, affixation of *-ity* is often accompanied by phonological changes in its base (*electric* ~ *electricity*) or with

7. There are several other universal preferences as well including figure and ground, and binarity. See Dressler and Kilani-Schoch (2016) for an explanation of these.

non-compositional meanings (*curiosity*, when it means ‘an odd thing’). In contrast, the meaning of a noun formed with *-ness* is almost always compositional, and *-ness* never has any phonological effect on its base. Thus, universally it is the more natural of the two morphological rules.

The preference for **biuniqueness** also has to do with the relationship between form and meaning. NM holds that languages prefer a one-to-one relationship between form and meaning, in other words that a single meaning should always be paired with a single form. Dressler and Kilani-Schoch (2016: 366) give as an example of biuniqueness the English classical combining form *-itis*, which when combined with a stem always means something having to do with ‘inflammation’ or more generally with ‘disease’, as in *tonsillitis* or *senioritis*.⁸ We can also assume that it follows from the preference for biuniqueness that agglutinative morphology should be considered more natural than fusional morphology, as agglutination favors a one-to-one relationship between form and meaning.

One of the interesting observations that NM makes is that these universal preferences are sometimes inconsistent with each other. For example, the preference for biuniqueness leads us to expect that agglutination is most natural, whereas the preference for indexicality leads us to expect a universal preference for fusional morphology. What do we make of these conflicts? Such conflicts actually lead us to the second level of naturalness that NM postulates, which is what NM calls Typological Naturalness. By Typological Naturalness, NM assumes that conflict between universal preferences may be decided in different ways in different languages, thus leading to typical clusters of typological characteristics. So languages that value biuniqueness and transparency over indexicality will be agglutinative in nature, and languages that value indexicality over other universal preferences will be fusional. Typological differences are thus made to follow from different rankings of universal preferences.

The final level recognized by NM is that of Language-Specific Naturalness, otherwise known as System Adequacy. According to Dressler and Kilani Schoch (2016: 371), this refers to “the identity of the language-specific morphological system – in other words, the typical or ‘normal’ structural properties of a system.” For example, in contemporary English, affixal inflection is the norm, and is favored over inflection effected by internal vowel change. So plurals with the suffix *-s* or past tenses with the suffix *-ed* are favored over plural or past tense forms like *mice* or *sang*. This means that in creating new plural or past tense forms (say, if we coined a new noun like *brouse* or a new verb like *zing*), it is preferred to stick with the affixal inflections (*brouses*, *zinged*), rather than creating new forms with internal vowel change (*brice*, *zang*).

8. Biuniqueness, we should point out, is actually fairly rare in morphology, where it is more the norm for meanings and forms to be in a many-to-many relationship. For example, the morpheme *-s* in English is associated with not only the plural of nouns but also the third person singular of verbs. And the meaning of ‘plural’ can also be expressed by the addition of *-en* in some forms or by vowel change in others (*mouse* ~ *mice*).

Although the functional orientation of NM is rather different than the formal focus of the other theories we have surveyed, it nevertheless shares some points of similarity with them. For example, like PM, NM has a typological outlook, seeking to understand how and why the languages of the world can differ from one another. Like CxM and PFM, NM accepts that complex words may be stored, but unlike those frameworks it recognizes both words and morphemes as units (Dressler and Kilani-Schoch 2016: 358). NM shares with CxM the belief that there should not be a strict separation between morphology and syntax, holding instead that the morphology–syntax boundary is fuzzy. For NM, there is a gradual continuum from syntax to morphology, although there are still principles that pertain specifically to morphology.

11.6 Naïve Discriminative Learning

There are a number of computational models in the current literature that share the leading idea that morphology is not a system of rules, schemas, or functions, but rather is an emergent phenomenon that arises as speakers learn successive new words. One prominent model of this sort is the framework called **Naïve Discriminative Learning (NDL)** (Baayen et al. 2011; Pirrelli, Plag, and Dressler 2020). Going back to our dichotomy between rules and storage, NDL assumes that both simple and complex words are learned as wholes and are stored, and that general mechanisms of learning (general in the sense that they account for all sorts of learning, not just linguistic learning) discern and extract regularities in those stored words. There are no rules of morphology per se, only patterns whose probability is weighed by our learning mechanisms. In frameworks of this sort “linguistic structure can emerge through self-organization of unstructured input ...” (Pirrelli, Plag, and Dressler 2020: 25). NDL aims to show how morphological patterns arise over the course of time by virtue only of the general mechanisms which allow humans to learn anything. As a computational model, NDL is based on a number of mathematical algorithms which I will leave aside here. Instead, I will try to present some of the main points of NDL in a general way.

NDL is basically a computer learning model that attempts to simulate lexical and morphological learning.⁹ Very roughly, in NDL words are conceived of as a series of **cues**, which are modeled as sequences of letters or sounds. These cues are presented to the learning model incrementally as sequences like those in (25). Each cue is a sequence of three successive letters/sounds, where # represents the start and end of the word; this series of cues is meant to simulate how a language learner might first hear or read the words *hands* and *farms*:

- (25) a. #ha, han, and, nds, ds#
 b. #fa, far, arm, rms, ms#

9. NDL is meant as a general linguistic model, so it can concern aspects of syntax and phonology as well as morphology and lexicon.

Each of these cues contributes to the model eventually recognizing the meaning *HAND + PLURAL* for the set of cues in (25a) and *FARM + PLURAL* for the set of cues in (25b); in NDL these ‘meanings’ are called **outcomes**.¹⁰ Each time one of these cues is encountered along with the outcome *HAND + PLURAL* OR *FARM + PLURAL*, that cue gains strength as a predictor of that meaning. One task of the learning model is to eventually discriminate that the *s* in the cues *nds* and *ds#* in (25a) or in the cues *rms* and *ms#* in (25b) is associated with the *PLURAL* meaning rather than the *HAND* OR *FARM* meanings. At the very beginning of the learning process, any one of the cues can theoretically signal any bit of the meaning. The association between *Xs#* (where *X* can be any letter/sound) and plural begins to emerge as the model encounters more and more cues in which both the final sequence of letters/sounds looks like *Xs#* and the outcome contains the meaning *PLURAL*. As more and more words are learned, the association between that *Xs#* and the meaning *PLURAL* gains strength and the association between the plural meaning and the other cues loses strength. The positive association that develops has the effect of a rule of pluralization without actually being a rule of pluralization. The strength of the association gives rise to a prediction of the model that the next time it encounters a string of cues that ends in *Xs#*, that string will have a high probability of being associated with the meaning *PLURAL*.

Of course, some of the sets of cues that the learning algorithm will encounter will contradict that generalization. Consider the cues in (26):

- (26) a. #ox, oxe, xen, en#
b. #bu, bus, us#

Suppose that the set of cues in (26a) is associated with the outcome *ox + PLURAL* and the set of cues in (26b) with the outcome *BUS*. In encountering these cues, the learning model will on the one hand find evidence that the meaning *PLURAL* can be associated with something other than the cue *Xs#* and on the other hand that the cue *Xs#* can sometimes be associated with something other than the meaning *PLURAL*. Each time the model encounters sets of cues like these, the probability that *Xs#* and plural are reliably associated will decrease a smidgeon. The point is, though, that the number of cues like those in (26) is small enough that the probability that *Xs#* is associated with plural will remain high enough to be a reliable predictor. Given a model like this, we can begin to study how long it takes for the computational model to reliably recognize the association between the *s* and the meaning *PLURAL*, at which point it can be said to have learned the plural.

The proponents of NDL argue that a major advantage of this model is that it is able to simulate aspects of processing of complex words, predicting reaction times¹¹ based on different influencing factors that have

10. The units of meaning in this model are called ‘lexomes’ (Pirrelli, Plag, and Dressler 2020: 45). In fact, lexomes are not exactly meanings, but the precise difference between the terms is not important in this very general introduction to NDL.

11. By reaction times, we mean the time it takes for a subject to recognize that a string of letters/sounds is or is not a word. See Chapter 2.

been studied in the psycholinguistic literature, for example, the frequency of various words or their family size (Baayen et al. 2011; Baayen, Hendrix, and Ramscar 2013; Milin et al. 2017). The argument that proponents of NDL make is that if the reaction times obtained in the computer simulation are similar to those obtained by subjects in actual psycholinguistic experiments, it provides evidence that what the computer does is similar to what humans do when they learn words.

Let's now compare NDL briefly to the other theoretical models we've looked at so far. With regard to the storage *versus* rules debate, NDL falls at the opposite end from DM. Both models seem to value simplicity and elegance, but in very different ways. DM achieves simplicity by doing without the storage of complex words and reducing morphology to syntactic rules. NDL achieves simplicity by using only a single general learning algorithm (no rules!) and assuming that all words are stored. With regard to units of analysis, like a number of the theories we have considered in this chapter, NDL does not recognize morphemes. Further, NDL is unlike most of the other theories that we have looked at in that it tries to model how lexical knowledge is acquired and organized over time and it attempts to account for the effects of frequency and family size on that learning. Although much of the research done in NDL has been focused on inflectional morphology, the model itself seems equally suited to analyzing either derivation or inflection.

11.7 The Lexical Semantic Framework

One of the things you might have noticed about most of the frameworks we've looked at so far is that when they talk about the meanings of complex words, they always use a sort of shorthand. For example, in Section 11.6, you saw that the NDL learning model had to learn to associate the meaning *ox* with some portion of the set of cues in (26a) or the meaning *HAND* with some portion of the set of cues in (25a). Or that in CxM, the meaning of the suffix *-ize* was represented as MAKE, CAUSE TO BE $X_{A,N}$. Or that DM relegated much of the semantic content of roots like $\sqrt{\text{DOG}}$ to something called the 'Encyclopedia'. The various theories that we've looked at so far have concentrated on aspects of the lexicon other than what simple and complex words mean. The actual meanings remain a sort of 'black box' – when we see root $\sqrt{\text{DOG}}$, in DM, we assume that *DOG* means something, but we don't really think about what that something looks like.

That 'something' is what the **Lexical Semantic Framework (LSF)**: Lieber 2004, 2016; Andreou 2021) is primarily concerned with; that is, LSF seeks to understand how speakers construct and understand the meanings of complex words. The central question that LSF tries to answer is why the relationship between form and meaning in complex words is rarely one-to-one. For example, as we saw in Chapters 3 and 10, there are affixes like *-er* in English that sometimes form agents (*writer*), sometimes instruments (*computer*), sometimes patients (*keeper*),

sometimes locations (*diner*), and so on; why does *-er* not mean just ‘agent’ and why can it express just this range of meanings? At the same time, there are actually several affixes in English besides *-er* that form agents (the *-ant* in *accountant*, or the *-ist* in *pianist*). Why do we have more than one affix to express this meaning? In fact, careful study of complex words in English reveals that most affixes have several meanings and most meanings are associated with several different affixes (Lieber 2016). LSF tries to delve into that black box of meaning I mentioned above, and in the process to explain why the many-to-many relationship between meaning and form is pervasive.

It does so by decomposing the semantic representations of bases and affixes alike into several parts. One part is called the **semantic skeleton**, which consists of a set of semantic features that form the core of meaning. The features represent fundamental, basic, and generally universal aspects of meaning, for example that a word like *dog* is a concrete (as opposed to abstract) and animate noun, represented by the features [+material] and [+animate], or that a word like *write* is an event or process represented by the feature [+dynamic]. The skeletons of simple words consist of these basic features plus one or more arguments. The argument that goes with the skeleton of *dog* tells us that *dog* is the kind of word that refers to a living thing; the subscript R stands for ‘referential’. The two arguments that go in the skeleton of *write* tell us (roughly) that *write* is an action word that takes a subject and an object:

- (27) a. *dog* [+material, +animate ([_R])]
 b. *write* [+dynamic([], [])]

Of course, the skeleton is not the complete semantic representation of *dog* or *write* – just the scaffolding for another part of the representation. The rest of the representation is what’s called the **semantic body** and consists of an assortment of information, possibly different from one speaker to the next, about material composition, perceptual characteristics, functions, purposes, origins, and so on. LSF presumes that English speakers all have the same skeletons for *dog* and *write*, but the body that one person has in their mental lexicon for *dog* might be different from another’s. For example, the body of the word *dog* for me includes lots about breeds, habits, coat color and texture, and training because I happen to be a dog person. Someone who’s never had a dog and rarely interacts with them might have a much skinnier body for the word *dog* than I do. On the other hand, I know that the word *aardwolf* must have the skeleton [+material, +animate ([_R])] just as *dog* does – an *aardwolf* is some sort of an animal – but the body of the word for me consists solely of the information that it’s wild (rather than domesticated) and maybe some sort of mammal.

Since LSF is a theory that seeks to explain the meanings of complex words, it must have a way of representing the meanings of affixes as well as of bases, and it must have a way of combining those meanings into more complex meanings. It does so by assuming that affixes have skeletons just as bases do. So the affix *-er* in English has the skeleton in (28)

- (28) Skeleton for the affix *-er* (Lieber 2016: 97):
 [+material, dynamic ([_R], <base>)]

What (28) says is that *-er* forms concrete nouns that have a processual flavor to them ([dynamic] usually signals an eventive meaning when it occurs in the skeleton of verbs); it has one argument which is referential, and it takes a base (that is, it's bound). Importantly in LSF, affixes generally have skeletons, but they don't necessarily have bodies. Often the skeleton is all there is.

When the skeleton of *write* is put together with the skeleton of the affix *-er*, we get the composite skeleton in (29):

- (29) [+material, dynamic ([_R,_i], [+dynamic ([_i], []))]
 -er *write*

The skeleton for the affix combines with that of the base and adds its meaning. In order to combine the two, they must come to refer to a single thing; the referential argument of the affix must latch onto or bind to one of the arguments of the base. In the case of *-er* it's the first argument of the verb, the one that usually refers to an agent, that gets bound, so *writer* comes to mean 'someone that writes'. If the affix were to attach to another verb, say, *compute* or *listen*, where the first argument refers to an instrument or an experiencer, we might get nouns like *computer* or *listener* that denote something other than agents.¹²

Remember that one of the central questions that LSF seeks to answer is why the relationship between form and meaning in derivation is rarely one-to-one. Because the meanings (that is, the skeletons) of affixes are very basic, they can give rise in combination with various kinds of bases to a range of meanings. In the example above, *-er* doesn't mean agent or instrument; it just means something concrete (a person or thing) with a processual flavor. The process of coindexation ensures that that concrete processual thing is generally associated with the subjects of verbs. On the other hand, the very sparseness of the system of features in LSF and the fact that affixes have no semantic bodies ensures that there are not so many possibilities for meanings of affixes. Because there are limited possibilities for affixal meanings in LSF, chances are that some meanings will be associated with more than one phonological form in a particular language.

Let us now compare LSF briefly to the other frameworks we have looked at. The first thing to note is that LSF has rather different goals than the other theories. Because all theories of the mental lexicon must occasionally be concerned with the meanings of words, theories like DM and CxM do make passing reference to lexical semantics. But they don't take lexical semantics as the specific object of study. Rather, they are more concerned with the overall architecture of the mental lexicon, that is, with form rather than meaning. Second, as we have seen, theories

12. LSF has a formal mechanism called the Principle of Coindexation that accounts for the combining of affixes with bases, but we won't go into the details here. See Lieber (2016: 94ff).

like PFM are primarily concerned with inflection rather than derivation. LSF, at least so far, has had almost nothing to say on the subject of inflectional meaning, concentrating instead on the sorts of meanings that we find in derivational word formation or compounding.

Finally, we might ask where LSF stands with respect to issues like the existence of morphemes or the storage *versus* rules question. Because LSF is a theory of meaning rather than form, it has remained relatively agnostic on those issues. LSF is compatible with formal theories that take the morpheme as a unit of analysis, but it could also be compatible with any theory that acknowledges that there is a difference in meaning between the word *writer* and the meaning *write*, whatever the formal relationship is between the words. Similarly, for LSF it doesn't matter whether words are formed by rules, stored as wholes, or both; whatever our conception of the mental lexicon is, we would still have to have some theory of lexical meaning.

11.8 A Brief Final Meditation on Theories

As this text has tried to show, there are many reasons for studying morphology and many ways of studying it. For some of you, it is enough to know what morphology is and to have a sense of how to approach a set of data from any language and make sense of it. However, the more time you spend looking at morphology and the more committed you become to studying it, the more you are likely to become an ardent proponent of one theory or another, or even to build your own theory. Grown-up morphologists often become quite partisan and argue that their theory is the best. And often, as in other aspects of linguistics (which is after all an empirical science), one theory does give a better account of some set of facts than another.

But what I have also tried to show in this brief chapter is that theories can start with different philosophical and psychological commitments and concentrate more on some aspects of morphology than on others, so that the overall theories are difficult to compare. There are issues that, say, DM is designed to study that PFM or CxM is not, and issues that NM or LSF considers central that are at best peripheral concerns in other theories. What I hope to have shown is that it's good to keep this in mind, and to remember that a theory that does not have the same empirical goals or philosophical commitments as your own is nevertheless worth reading about. The best thing I can do at the end then is to leave you with more to read!

Further Reading

General

Audring, Jenny and Francesca Masini. 2019. *The Oxford Handbook of Morphological Theory*. Oxford University Press.

Distributed Morphology (DM)

Halle, Morris and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale and Samuel J. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 111–76. Cambridge, MA: MIT Press.

Siddiqi, Daniel. 2019. Distributed Morphology. In Jenny Audring and Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 143–65. Oxford University Press.

Construction Morphology (CxM)

Booij, Geert. 2010. *Construction Morphology*. Oxford University Press.

Masini, Francesca and Jenny Audring. 2019. Construction Morphology. In Jenny Audring and Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 365–89. Oxford University Press.

Paradigm Function Morphology (PFM)

Bonami, Olivier and Gregory Stump. 2016. Paradigm Function Morphology. In Andrew Hippisley and Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 449–81. Cambridge University Press.

Stump, Gregory. 2001. *Inflectional Morphology*. Cambridge University Press.

Natural Morphology (NM)

Dressler, Wolfgang. 1985. *Morphonology*. Ann Arbor, MI: Karoma Press.

Dressler, Wolfgang and Marianne Kolani-Schoch. 2016. Natural Morphology. In Andrew Hippisley and Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 356–89. Cambridge University Press.

Naïve Discriminative Learning (NDL)

Baayen, Harald, Petar Milin, Dusica Đurđević, Peter Hendrix, and Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naïve discriminative learning. *Psychological Review* 118 (3): 438–81.

Pirrelli, Vito, Claudia Marzi, Marcello Ferro, Franco Alberto Cardillo, Harald R. Baayen, and Petar Milin. 2020. Psycho-computational modelling of the mental lexicon: a discriminative learning perspective. In Vito Pirrelli, Ingo Plag, and Wolfgang Dressler (eds.), *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon*, 23–82. Berlin: De Gruyter. DOI: [10.1515/9783110440577](https://doi.org/10.1515/9783110440577)

Lexical Semantic Framework (LSF)

Andreou, Marios. 2021. Lexical Semantic Framework for Morphology. In Rochelle Lieber (ed.), *The Oxford Encyclopedia of Morphology*, 1017–31. Oxford: Oxford University Press.

Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge University Press.

Glossary

- ablative:** The case typically assigned to objects of prepositions denoting instruments or sources.
- ablaut:** Internal vowel change. Also known as apophony.
- absolutive:** In an ergative-absolutive case system, the case that is assigned to the subject of an intransitive clause and the object of a transitive clause.
- accusative:** In a nominative-accusative case system, the case assigned to the direct object of the clause, and in some languages to objects of prepositions.
- acronym:** A word made up of the initial letter or letters of a phrase and pronounced as a word. For example, from *self-contained underwater breathing apparatus* we get the acronym *scuba*, pronounced [skubə].
- active:** A voice in which the subject of the clause is (typically) the agent, instrument, or experiencer and the direct object the theme or patient. In English an active clause would be *Fenster ate the pizza*, as opposed to a passive *The pizza was eaten*.
- adjuncts:** Non-argumental phrases that are not necessary to the meaning of a verb.
- affix:** A bound morpheme that consists of one or more segments that typically appear before, after, or within a base morpheme.
- affixal polysemy:** Multiple related meanings of an affix.
- affixation:** Formation of words by the addition of prefixes, suffixes, infixes, and circumfixes.
- agent:** The argument of the verb that performs or does the action. Agents typically are sentient and have intentional or volitional control of actions.
- agglutinative:** One of the four traditional classifications of morphological systems. Agglutinative systems are characterized by sequences of affixes each of which is easily segmentable from the base and associated with a single meaning or grammatical function.
- agrammatism:** A form of aphasia in which comprehension is good, production is labored, and grammatical or function words largely absent. Also known as non-fluent aphasia.
- agreement:** Contextual inflection of elements of a phrase or sentence to match another element of that phrase or sentence. For example, in the Romance languages the inflection of adjectives in a noun phrase must match the gender and number of the head noun. In Latin the verb must be inflected to match the person and number of its subject.
- allomorph:** A phonologically distinct variant of a morpheme.
- analytic:** See [isolating](#).
- anti-passive:** Morphology that decreases the valency of verbs by eliminating the object argument.

- aphasia:** A language disorder that results from injury to the brain.
- apophony:** Internal vowel change. Also known as ablaut.
- applicative:** Morphology that increases the valency of a verb by adding an object argument.
- argument:** A noun phrase that is semantically and often syntactically necessary to the meaning of a verb. The arguments of a verb consist of its subject and complement(s).
- aspect:** A type of inflection that conveys information about the internal composition of an event.
- assimilation:** A phonological process in which segments come to be more like each other in some phonological feature such as voicing or nasality.
- attenuative affixes:** Affixes that denote 'sort of X' or 'a little X'.
- attributive compound:** A compound in which the two elements bear a modifier–modified relationship to one another.
- augmentative:** A kind of expressive morphology which conveys notions of larger size and sometimes pejorative tone.
- backformation:** A morphological process in which a word is formed by subtracting a piece, usually an affix, from a word which is or appears to be complex. In English, for example, the verb *peddle* was created by backformation from *peddler* (originally spelled *peddlar*).
- base-driven selection:** Choice of an affix by its base, whether a simple or complex word. For example, in English, words prefixed by *en*-always form nouns by suffixation of *-ment*. The complex base *enX* therefore selects its affix.
- binyan:** A templatic pattern associated with a specific meaning or function.
- biuniqueness:** In the framework of Natural Morphology, the principle that the relationship between form and meaning should always be one-to-one.
- blending:** A type of word formation in which parts of words that are not themselves morphemes are combined to form a new word. For example, the word *smog* is a blend of *smoke* and *fog*.
- blocking:** The tendency of an already existent word to preclude the derivation of another word that would have the same meaning. For example, the existence of the word *glory* precludes the derivation of *gloriosity* and the existence of *went* precludes the formation of the regular past tense *goed*.
- bound base:** A morpheme which is not an affix but which nevertheless cannot stand on its own. In English, bound bases are items like *endo*, *derm*, and *ology*, from which neo-classical compounds like *endoderm* and *dermatology* are formed.
- bracketing paradoxes:** Complex words in which there is a mismatch between syntactic structure and phonological form or between syntactic structure and semantic interpretation. Within theories that admit stratal ordering, bracketing paradoxes can also involve mismatches between the structure required on the basis of word formation rules and the structure consistent with stratal ordering.
- case:** Inflectional marking which signals the function of noun phrases in sentences.
- causative:** Valency-changing morphology that adds an external causer to a verb.

circumfix:	A morpheme that consists of the simultaneous attachment of a prefix and a suffix which convey meaning or function only when they appear together.
classifier:	A word or morpheme used in a noun phrase, often in the context of a numeral or determiner, that designates the grammatical or semantic class of the noun.
clipping:	A word formed by subtraction of part of a larger word. For example, in English <i>math</i> is a clipping from <i>mathematics</i> and <i>ad</i> is a clipping from <i>advertisement</i> .
clitic:	Small grammatical elements that cannot occur independently but are not as closely bound to their hosts as inflectional affixes are.
closed class:	A fixed list from which particular forms can be lost, but to which no new forms can be added.
coda:	The consonants that come after the nucleus in a syllable.
coinage:	A word that is made up from whole cloth rather than by affixation, compounding, conversion, blending, reduplication, or other processes.
competition:	When rival affixes interact such that a derived word can be formed with one of the set, precluding the others from forming synonymous words.
completive:	An aspectual distinction that focuses on the end of an event.
complex word:	A word made up of more than one morpheme.
Complexity Based Ordering:	The hypothesis that suffixes which are more transparent, more productive, and more easily segmented from their bases will occur outside those that are less transparent, less productive, and less easily segmented from their bases.
compositional:	The semantic interpretation of a word is compositional to the extent that it can be computed as the sum of the meanings of each of its morphemes.
compound:	A word made up of two or more separate lexemes.
compounding:	Formation of words by combining bases.
conjugation:	The traditional name for the inflectional paradigm of a verb.
consonant mutation:	A form of internal stem change in which consonants of a base differ systematically in different morphological contexts.
construction:	A grammatical structure or word formation pattern that imparts meaning apart from the words or morphemes that make it up.
Construction Morphology (CxM):	A theory of morphology that holds that our mental lexicon is comprised of constructions, that is, pairings of complex words and their meanings.
contextual inflection:	Inflection which is determined by the syntactic construction in which a word finds itself.
continuative:	An aspectual distinction that focuses on the middle of an event as it progresses.
conversion:	A type of word formation in which the category of a base is changed with no corresponding change in its form. For example, in English

- the verb *to chair* is formed by conversion from the noun *chair*. Also called functional shift.
- coordinative compound:** A type of compound in which the two elements have equal semantic weight. Examples in English are *producer-director* or *blue-green*.
- corpus:** A database comprised of spoken language and/or written texts that can be mined for various forms of linguistic study.
- cran morph:** A bound morpheme that occurs in only one word. An example in English is *cran* in *cranberry*.
- creativity:** The conscious use of unproductive word formation processes to form new words that are often perceived as humorous, annoying, or otherwise worthy of note.
- cue:** In the framework of Naïve Discriminative Learning, sequences of letters or sounds that are presented to the learning model.
- cumulative exponence:** The association of several inflectional meanings with a single inflectional morpheme.
- dative:** In languages which mark case, the case assigned to the indirect object and frequently to objects of prepositions.
- declarative:** The mood/modality of ordinary statements (as opposed to questions or imperatives, for example).
- declension:** The traditional name for the inflectional paradigm of a noun, especially in languages that display case marking.
- default endings:** Inflectional markings that are used when no more specific marking is applicable.
- defectiveness:** Cells (combinations of inflectional features) in paradigms for which no form exists.
- dependent:** Modifiers and complements of the head of a phrase.
- dependent-marking:** Morphological marking of the dependents of a phrase rather than its head. For example, in noun phrases marking occurs on determiners and adjectives rather than the noun.
- derivation:** Lexeme formation processes that either change syntactic category or add substantial meaning or both.
- Developmental**
- Language Disorder:** See [Specific Language Impairment](#).
- deverbal compound:** See [synthetic compound](#).
- diminutive:** Evaluative morphology that expresses smallness, youth, and/or affection.
- Distributed**
- Morphology (DM):** A theoretical model that holds that morphology is not to be distinguished from syntax. Complex words are derived by rules like Merge and Move.
- ditransitive:** The valency of a verb that has two complements.
- double marking:** Morphological marking of both the head of a phrase and its dependents. For example, in a noun phrase marking would occur on both the head noun and on adjectives and/or determiners that modify it.
- dual:** Number marking that denotes exactly two objects.
- electroencephalography (EEG):** A device for measuring electrical activity on the scalp.

enclitic:	A clitic that is positioned after its host.
endocentric:	Having a head. In endocentric compounds the compound as a whole is the same category and semantic type as its head.
epenthesis:	A phonological process in which a segment is inserted between two other segments.
ergative:	In an ergative–absolutive case system, the marking of the subject of a transitive verb.
ergative–absolutive case system:	A case-marking system in which the subject of an intransitive verb is marked with the same case as the object of a transitive verb, and the subject of a transitive verb receives a different marking.
etymology:	The study of the origins and development of words.
evaluative affixes:	Affixes, including diminutives and augmentatives, that denote size and/or negative or positive associations.
event related potential (ERP):	A neuroimaging technique that measures electrical activity in neurons and can track the brain's reaction to a linguistic stimulus over the course of time.
evidentiality:	An inflectional category in which speakers must specify what the source of knowledge is.
exclusive:	Person-marking in which the hearer is not included.
exocentric:	Lacking a head. In exocentric compounds the compound as a whole is not of the category or semantic type of either of its elements.
experiencer:	The argument of the verb that experiences the action. Like agents, experiencers are sentient, but unlike agents they are not in volitional control of the action. They are often the subjects of verbs of perception or emotion.
extender:	An extra segment that occasionally appears in derived words in English, but which has no meaning and does not obviously belong to either the base or the affix, for example, <i>n</i> in <i>Platonic</i> (<i>Plato</i> + <i>n</i> + <i>ic</i>).
eye tracking:	An experimental method in which the movements of a subject's eyes are followed and recorded while they are reading words or sentences.
fast mapping:	The ability of language learners to rapidly create lexical entries for new words that they hear.
formative:	An element that is used to form words but which lacks clear meaning or function.
free base:	A base that can occur as an independent word.
frequency of base type:	The number of different bases that are available for an affix to attach to, thus resulting in new words.
frequentative:	Aspectual marking that signals repetition of an action. See also iterative .
full reduplication:	A word formation process in which whole words are repeated to denote some inflectional or derivational meaning.
functional magnetic resonance imaging (fMRI):	A device which maps brain activity over the course of time.
functional shift:	See conversion .

- fusional:** One of the four traditional classifications of morphological systems. In fusional systems words are complex but not easily segmentable into distinct morphemes. Morphological markings may bear more than one function or meaning.
- future:** Tense that signals that an event will occur subsequent to the point of speaking.
- Gavagai problem:** A philosophical problem concerning how children come to associate the meaning of a word with the action or entity the word denotes.
- gender:** Inflectional classes of noun that may be either arbitrary (**grammatical gender**) or semantically based (**natural gender**). See also **noun classes**.
- genitive:** The case assigned to the possessor of a noun.
- gloss:** A morpheme-by-morpheme translation of a word or sentence.
- habilitative:** A verb form meaning 'can V'.
- habitual:** Aspectual marking that designates that an action is usually or characteristically done.
- hapax legomenon:** A word that occurs only once in a corpus.
- head:** The morpheme that determines the category and semantic type of the word or phrase.
- head-marking:** Morphological marking of the head of a phrase rather than its dependents. For example, in noun phrases marking occurs on the noun itself, rather than on determiners and adjectives that modify the noun.
- heavy syllable:** A syllable which has either a coda or a long vowel or both.
- hypocoristic:** A nickname.
- iconicity:** In Natural Morphology, a relationship in which form and meaning are similar or analogous in some way.
- imperative:** The mood/modality used for commands.
- imperfective:** Aspectual distinction in which the event is viewed from inside as ongoing.
- implicational universal:** In linguistic typology a generalization that if one linguistic characteristic is found in a language, another characteristic is expected to occur as well.
- inceptive:** Aspectual distinction that focuses on the beginning of an event.
- inclusive:** Person-marking that includes the hearer as well as the speaker.
- index of exponence:** Typological measure of how many meanings may be packed into a single inflectional morpheme in a language.
- index of fusion:** Typological measure of how easily separable morphemes are in a language.
- index of synthesis:** Typological measure of how many morphemes there are per word in a language.
- indexicality:** In Natural Morphology, the universal preference for having a morpheme that modifies a base be as close as possible to that base.
- infix:** An affix which is inserted into a base morpheme, rather than occurring at the beginning or the end.
- inflection:** Word formation process that expresses a grammatical distinction.
- inflectional class:** Different inflectional subpatterns displayed by a category. See also **noun classes**, **gender**.

- inherent inflection:** Inflection that does not depend on context. For example, the inflectional category of aspect is inherent in verbs. The inflectional category of number is inherent in nouns.
- initialism:** A word created from the first letters of a phrase, and pronounced as a sequence of letters. For example, FBI is an initialism created from Federal Bureau of Investigation, and pronounced [ɛf bi ai].
- instrument:** The argument of the verb that is used to perform an action.
- intensive:** Morphological form that means ‘very X’ or ‘X a lot’.
- interfix:** See [linking element](#).
- internal stem change:** Morphological process which changes a vowel or consonant in the stem. Also sometimes called **simulfixation**. Internal vowel change is called ablaut and internal consonant change is called consonant mutation.
- interrogative:** The mood/modality of questions.
- intransitive:** The valency of a verb that takes only one argument.
- irrealis:** A mood/modality signaling that an event is imagined or thought of but not verifiable.
- isolating:** One of the traditional four classifications of morphological systems. In isolating systems words consist of only one morpheme. Also known as [analytic](#).
- Item and Arrangement Model (IA):** A theoretical model of word formation in which affixes have lexical entries just as bases do, and words are built by rules which combine bases and affixes hierarchically.
- Item and Process Model (IP):** A theoretical model of word formation in which derivation and inflection are accomplished by rules that add affixes, or perform reduplication, internal stem change, and other processes of word formation.
- iterative:** Aspectual distinction that signals that an action is done repeatedly. See also [frequentative](#).
- Late Insertion:** In Distributed Morphology, the idea that the phonetic reflexes of bundles of inflectional features are inserted only after all syntactic processes have applied.
- learning mechanism:** In Naïve Discriminative Learning, an algorithm for extracting patterns from data to which the learning model has been exposed.
- lexeme:** Families of words that differ only in their grammatical endings or grammatical forms. For example, the words *walk*, *walking*, *walked*, and *walks* all belong to the same lexeme.
- Lexical Contrast Principle:** The principle that the language learner will always assume that a new word refers to something that does not already have a name.
- lexical decision task:** A kind of experiment in which subjects must decide whether a word they are exposed to is an existing word or not.
- Lexical Integrity Hypothesis:** The hypothesis that syntactic rules may not create or affect the internal structure of words.

Lexical Semantic Framework (LSF):	A theoretical framework that seeks to model the semantics of both simple and complex words.
lexical strata:	Layers of word formation within a single language that display different phonological properties and different patterns of attachment.
lexicalization:	The process by which complex words come to have meanings that are not compositional.
lexicalized:	The property of a word's having a meaning that is not the sum of the meanings of its parts.
lexicographer:	One who writes dictionaries.
light syllable:	A syllable that has only a short vowel and no coda.
linking element:	A meaningless vowel or consonant that occurs between the two elements that make up a compound.
logographic writing:	A writing system in which each symbol stands for one word.
magnetoencephalography (MEG):	A device which measures magnetic activity in the brain.
mental lexicon:	The sum total of all the information a native speaker of a language has about the words, morphemes, and morphological rules of her/his language.
mirativity:	A category of inflection that marks information as surprising or unexpected.
mood/modality:	Inflectional distinctions that signal the kind of speech act in which a verb is deployed.
morpheme:	The smallest meaningful part of a word.
multiple exponence:	The property of having an inflectional distinction marked in a single word by more than one morpheme.
Mutual Exclusivity Principle:	The tendency of language learners to assume that each object has one and only one name.
Naïve Discriminative Learning (NDL):	A computational model of morphology that seeks to analyze inflection and word formation as an emergent phenomenon that arises as speakers learn successive new words.
Natural Morphology (NM):	A theoretical framework that seeks to ground the study of morphology in general cognitive and semiotic forces that drive language to be the way it is.
naturalness:	In Natural Morphology, qualities of language that are cross-linguistically frequent, most easily acquired by children, most easily processed, most resistant to language change, and least affected by disease or trauma that causes damage to the language faculty.
negative affix:	An affix that means 'not-X'.
neo-classical compound:	In English, a compound that consists of bound bases that are derived from Greek or Latin.
nominative:	In a nominative–accusative case system, the case assigned to the subject of the sentence.

nominative–accusative case system:	A case system in which the subject of a transitive sentence receives the same marking as the subject of an intransitive sentence, and the object of a transitive sentence receives a different case.
nonce word:	A word that occurs only once.
noun classes:	Groupings of nouns that share the particular inflectional forms that they select for. Noun classes can be based roughly on gender, shape, animacy, or some combination of these semantic properties, but frequently the membership in noun classes is largely arbitrary.
noun incorporation:	A form of word formation in which a single compound-like word consists of a verb or verb stem and a noun or noun stem that functions as one of its arguments, typically its object.
nucleus:	The vowel in a syllable.
number:	An inflectional distinction that marks how many entities there are.
onset:	The consonants that come before the nucleus in a syllable.
orthography:	The spelling system of a language.
outcome:	In Naïve Discriminative Learning, those characteristics, for example meanings, that come to be associated with cues.
overabundance:	A characteristic of inflection in which more than one form is associated with a single cell in a paradigm.
palatalization:	A phonological process by which one segment takes on a palatal point of articulation, frequently in the environment of a front vowel.
paradigm:	A grid or table consisting of all of the different inflectional forms of a particular lexeme or class of lexemes.
Paradigm Function:	In Paradigm Function Morphology, a type of rule that associates a root plus its morphosyntactic properties with a word form.
Paradigm Function Morphology (PFM):	A theoretical framework that seeks to characterize inflectional systems as associations between morphosyntactic features and word forms through their organization into paradigms.
parasyntesis:	A type of word formation in which a particular morphological category is signaled by the simultaneous presence of two morphemes.
partial reduplication:	A type of word formation in which part of a base morpheme is repeated.
participant affix:	See personal affix .
passive:	A voice in which the theme/patient of the verb serves as the subject and the agent is either absent or marked by a preposition or oblique case marking.
past:	Tense that signals that an action has occurred before the time of the speaker's utterance.
patient:	The noun phrase in a sentence that undergoes the action.
perfect:	An aspectual distinction that expresses something that happened in the past but still has relevance to the present.
perfective:	An aspect in which an event is viewed as completed. The event is viewed from the outside, and its internal structure is not relevant.
periphrastic marking:	Marking by means of separate words, as opposed to morphological processes. For example, in English one- or two-syllable adjectives form

	the comparative by affixation of <i>-er</i> (<i>redder</i> , <i>happier</i>) but three-syllable adjectives form their comparatives periphrastically (<i>more intelligent</i>).
person:	Inflectional distinction that expresses the involvement of the speaker, the hearer, or a person other than the speaker or hearer.
personal affix:	Derivational affixes that produce either agent nouns (<i>writer</i> , <i>accountant</i>) or patient nouns referring to humans (<i>employee</i>).
phrasal compound:	A compound that consists of a phrase or sentence as its first element and a noun as its second element. For example, <i>stuff-blowing-up effects</i> .
phrasal verb:	A combination of a verb plus a preposition, frequently having an idiomatic meaning. Phrasal verbs have the characteristic that the preposition can and sometimes must occur separated from its verb. For example, <i>call up</i> .
place assimilation:	A phonological process in which a consonant assimilates to the place of articulation of a preceding or following consonant.
polysynthetic:	One of the four traditional typological classifications of morphological systems. In polysynthetic languages words are frequently extremely complex, consisting of many morphemes, some of which have meanings that are typically expressed by separate lexemes in other languages.
positron emission tomography (PET) scan:	An imaging technique that measures the level of blood flow to different parts of the brain, which in turn shows us areas of activation in those parts.
prefix:	An affix that goes before the base.
prepositional/relational affix:	Affixes that convey notions of space and time. For example, <i>over-</i> , <i>pre-</i> .
present:	Tense relating the speaker's utterance to the moment of speaking.
primary compound:	See root compound .
privative affixes:	Affixes that denote 'without X' (e.g., <i>-less</i> in English) or 'remove X' (e.g., <i>de-</i> in English).
proclitic:	A clitic that is positioned before its host.
productivity:	The extent to which a morphological process can be used to create new words.
progressive:	Aspectual distinction that expresses ongoing action.
quantificational:	An aspect denoting the number of times or the frequency with which an action is done.
quantitative affixes:	Affixes that express something relating to amount (e.g., <i>multi-</i> or <i>-ful</i> in English).
reaction time:	In a lexical decision experiment, the amount of time it takes a subject to decide whether the stimulus they are exposed to is a real word or not.
realis:	A mood/modality in which the speaker means to signal that the event is actual, that it has happened or is happening, or is directly verifiable by perception.
realizational model:	A theoretical model of word formation that does not separate out morphemes into discrete pieces, but rather states rules that associate meanings (single or multiple) with complex forms.

Realization Rules:	In Paradigm Function Morphology, rules that express generalizations over Paradigm Functions. These include Rules of Exponence and Rules of Referral.
reduplication:	A morphological process whereby words are formed by repeating all or part of their base.
response latency:	See reaction time .
rhyme:	A constituent in a syllable structure made up of the nucleus plus the coda, if there is one.
Righthand Head Rule:	A theoretical hypothesis that defines the head of a morphologically complex word to be the righthand member of that word.
rival affixes:	A group of affixes that have the same meaning or function.
root:	The part of a word that is left after all affixes have been removed. Roots may be free bases, as is frequently the case in English, or bound morphemes, as is the case in Latin.
root and pattern morphology:	See templatic morphology .
root compound:	A compound in which the head element is not derived from a verb (cf. synthetic compound). <i>Dog bed</i> , <i>windmill</i> , <i>blue-green</i> , and <i>stir-fry</i> are root compounds. Also known as primary compound.
Rules of Exponence:	In Paradigm Function Morphology, a type of realization rule that associates roots that have specific morphosyntactic features with the phonological material that marks those features.
Rules of Referral:	In Paradigm Function Morphology, a type of realization rule that systematically associates the exponence of one morphological form with that of another.
schema:	In Construction Morphology, a generalization over forms in our mental lexicons that can then also serve to generate new forms.
semantic body:	In the Lexical Semantic Framework, those aspects of lexical meaning that are encyclopedic and that need not be the same from one speaker to another.
semantic skeleton:	In the Lexical Semantic Framework, the core aspects of lexical meaning that are expressed as semantic features and their arguments.
semelfactive:	An aspectual distinction that expresses that an action is done just once.
separable prefix verb:	A kind of verb found in Dutch and German which consists of two parts which frequently together have an idiomatic meaning and which occur as one word in some syntactic contexts but separated from each other in other syntactic contexts.
simple clitic:	A clitic that appears in the same position as the independent word of which it is a variant. In English, the contractions <i>'ll</i> and <i>'d</i> are simple clitics.
simplex:	Consisting of one morpheme.
simulfix:	See internal stem change .
sound symbolism:	A sequence of sounds that is sometimes associated with a vague meaning but which does not have the status of a morpheme. For example, <i>sn</i> in <i>sneeze</i> , <i>snort</i> , <i>sniffle</i> .
special clitic:	A clitic that is not a reduced form of an independent word. The object pronouns in Romance languages are examples of special clitics.

Specific Language Impairment (SLI):	A genetic disorder in which individuals display normal intelligence and have no hearing impairment, but are slow to produce and understand language and display speech characterized by the omission of various inflectional morphemes.
speech acts:	Ways in which we can use words to perform actions, for example, asking a question or giving a command.
Spell Out:	In Distributed Morphology, the derivational stage where vocabulary items are associated with bundles of morphosyntactic features that have been generated by syntactic rules.
splinter:	A piece of a word that is not itself a morpheme but which can be used on an analogical basis to form a limited set of new words, for example, <i>-tarian</i> from <i>vegetarian</i> as it is used to form <i>flexitarian</i> .
stem:	The part of a word that is left when all inflectional endings are removed.
Stratal Ordering Hypothesis:	The hypothesis that English morphology is divided into levels, each of which is comprised of a set of affixes and phonological rules. Strata are strictly ordered with respect to each other such that the rules of an earlier stratum cannot apply to the output of a later stratum.
strong verb:	In Germanic languages, verbs whose past tenses and past participles are formed by internal stem change.
subjunctive:	A mood/modality that is used to express counterfactual situations or situations expressing desire.
subordinative compound:	A compound in which one element bears an argumental relation to the other. Compounds like <i>truck driver</i> or <i>dog attack</i> in English are subordinative.
subtractive morphology:	A morphological process which deletes part of its base, rather than adding to or changing it.
suffix:	An affix that goes after the base.
suppletion:	An instance in which one or more of the inflected forms of a lexeme are built on a base that bears no relationship to the base of other members of the paradigm.
syllable:	A unit of linguistic structure consisting of a vowel (see nucleus) possibly with preceding consonants (see onset) and/or following consonants (see coda).
syncretism:	An instance in which two or more cells in a paradigm are filled with the same form.
synthetic compound:	A compound in which the head is derived from a verb and the non-head bears an argumental relationship to the head. Examples of synthetic compounds in English are <i>truck driver</i> and <i>hand washing</i> .
template:	In a root and pattern system of morphology, a pattern of consonants and vowels that is associated with some meaning.
templatic morphology:	A kind of morphological process in which words are derived by means of arranging morphemes according to meaningful patterns

	of consonants and vowels or templates . Also called root and pattern morphology, simulfixation , or transfixation .
tense:	Inflectional morphology that gives information about the time of an action.
theme:	The noun phrase in a sentence that gets moved by the action.
theme vowel:	In languages like Latin and the Romance languages, the vowel that attaches to the root before inflectional and derivational affixes are added.
token:	In counting words in a text or corpus, each instance of a word counts as a token of that word. This gives the raw number of words that occur with a particular affix.
transfix:	See templatic morphology .
transitive:	A valency in which a verb takes two arguments, generally a subject and object.
transparency:	In Natural Morphology, the characteristic that meanings and phonological forms of morphemes in complex words are typically compositional and easily separable.
transparent process:	A morphological process resulting in words that can be easily segmented such that there is a one-to-one correspondence between form and meaning.
transpositional affixes:	Affixes that change syntactic category without adding meaning.
triliteral root:	A root consisting of three consonants. These typically occur in the templatic morphology of the Semitic languages.
type:	In counting words in a text or corpus, only the first instance of each word is counted. This gives the number of types with a particular affix.
typology:	Linguistic subfield that attempts to classify languages according to kinds of structures, and to find correlations between structures and genetic or areal characteristics.
umlaut:	Phonological process in which the vowel of the base is fronted or raised under the influence of a high vowel in the following syllable.
underlying representation:	The basic mental representation of a morpheme from which various allomorphs can be derived via phonological rule.
Unitary Base Hypothesis:	The theoretical hypothesis that affixes will not select bases of more than one category.
usefulness:	The extent to which a morphological process produces words that are needed by speakers.
valency:	The number of arguments selected by a verb.
Vocabulary Item:	In Distributed Morphology, the phonological representation that comes to be associated with a bundle of morphosyntactic features.
voice:	A category of inflection that allows different arguments to be focused in sentences. In active voice sentences, the agent is typically focused because it is the subject, and in passive sentences, the patient is focused because it is the subject.

- voicing assimilation:** A phonological process whereby segments come to be voiced in the environment of voiced segments or voiceless in the environment of voiceless segments.
- vowel harmony:** A phonological process whereby all the vowels of a word come to agree in some phonological feature, for example in backness or rounding.
- weak verb:** In the Germanic languages, verbs that form their past tenses and participles by suffixation.
- Whole Object Principle:** The principle that word learners will not assume that a new word refers to a part of the object or its color or shape if they do not already have a word for the object as a whole.
- Williams Syndrome:** A genetic disorder in which individuals (in addition to certain physical traits and some developmental delay) speak fluently and produce sentences with correct regular past tenses, but have more trouble with irregular ones.
- word:** A linguistic unit made up of one or more morphemes that can stand alone in a language.
- Word and Paradigm (WP) Model:** See [realizational model](#).
- word forms:** Differently inflected forms that belong to the same lexeme. For example, *walks*, *walking*, *walk*, and *walked* are all word forms that belong to the same lexeme.
- zero affixation:** An analysis of conversion in which a change of part of speech or semantic category is effected by a phonologically null affix.

References

- Aikhenvald, Alexandra. 2000. Unusual classifiers in Tariana. In Gunter Senft (ed.), *Systems of Nominal Classification*, 93–113. Cambridge University Press.
2004. *Evidentiality*. Oxford University Press.
- Alford, Henry. 2005. Not a word. *The New Yorker* 81, 25: 32.
- Amith, Jonathan D. and Thomas C. Smith-Stark. 1994. Transitive nouns and split possessive paradigms in Central Guerrero Nahuatl. *International Journal of American Linguistics* 60, 4: 342–68.
- Anderson, Stephen. 1982. Where's morphology? *Linguistic Inquiry* 13: 571–612.
1992. *A-Morphous Morphology*. Cambridge University Press.
2005. *Aspects of the Theory of Clitics*. Oxford University Press.
- Andreou, Marios. 2021. Lexical Semantic Framework for morphology. In Rochelle Lieber (ed.), *The Oxford Encyclopedia of Morphology*, 1017–31. Oxford University Press.
- Arnott, D. W. 1970. *The Nominal and Verbal Systems of Fula*. Oxford University Press.
- Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.
- Audring, Jenny and Francesca Masini. 2019. *The Oxford Handbook of Morphological Theory*. Oxford University Press.
- Baayen, Harald. 1989. A Corpus-based Approach to Morphological Productivity: Statistical Analysis and Psycholinguistic Interpretation. Dissertation, Free University, Amsterdam.
2014. Experimental and psycholinguistic approaches. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 95–117. Oxford University Press.
- Baayen, Harald and Rochelle Lieber. 1991. Productivity and English derivation: a corpus based study. *Linguistics* 29: 801–43.
- Baayen, Harald, Petar Milin, Dusica Đurđević, Peter Hendrix, and Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naïve discriminative learning. *Psychological Review*. 118 (3): 438–481.
- Baayen, Harald, Peter Hendrix, and Michael Ramscar. 2013. Sidestepping the combinatorial explosion: an explanation of n-gram frequency effects based on naïve discriminative learning. *Language and Speech* 56(3): 329–347.
- Badecker, William and Alfonso Caramazza. 1999. Morphology and aphasia. In Andrew Spencer and Arnold Zwicky (eds.), *The Handbook of Morphology*, 390–405. Oxford: Blackwell.
- Baerman, Matthew. 2007. Syncretism. *Language and Linguistics Compass* 1, 4: 539–51.
- Baker, Mark. 1988. *Incorporation*. University of Chicago Press.
1996. *The Polysynthesis Parameter*. Oxford University Press.
- Baker, Mark and Carlos Fasola. 2009. Araucanian: Mapudungun. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 594–608. Oxford University Press.
- Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge University Press.
2003. *Introducing Linguistic Morphology*. 2nd edition. Washington, DC: Georgetown University Press.
2014. Concatenative derivation. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 118–35. Oxford University Press.
- Bauer, Laurie, Rochelle Lieber, and Ingo Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford University Press.
- Bauer, Winifred. 1993. *Maori*. London: Routledge.
- Baugh, Albert and Thomas Cable. 2013. *A History of the English Language*. 6th edition. London: Routledge.

- Berman, Ruth. 2009. Children's acquisition of compound structures. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 272–97. Oxford University Press.
- Bhat, D. N. S. 1999. *The Prominence of Tense, Aspect and Mood*. Amsterdam: John Benjamins.
- Bisetto, Antonietta and Sergio Scalise. 2005. The classification of compounds. *Lingue e Linguaggio* 4, 2: 319–32.
- Bishop, Dorothy. 2017. Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language and Communication Disorders*. 52, 6: 671–80. DOI: 10.1111/1460-6984.12335.
- Blevins, Juliette. 1999. Untangling Leti infixation. *Oceanic Linguistics* 38, 2: 383–403.
2014. Infixation. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 136–53. Oxford University Press.
- Bloom, Paul. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.
- Bonami, Olivier and Gregory Stump. 2016. Paradigm Function Morphology. In Andrew Hippisley and Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 449–81. Cambridge University Press.
- Booij, Geert. 2002. *The Morphology of Dutch*. Oxford University Press.
2010. *Construction Morphology*. Oxford University Press.
2016. Construction Morphology. In Andrew Hippisley and Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 424–48. Cambridge University Press.
- Booij, Geert and Rochelle Lieber. 2004. On the paradigmatic nature of affixal semantics in English and Dutch. *Linguistics* 42: 327–57.
- Borer, Hagit. 2013. *Taking Form*. Oxford University Press.
- Brown, Roger. 1973. *A First Language*. Cambridge, MA: Harvard University Press.
- Budd, Mary-Jane, Silke Paulmann, Christopher Barry, and Harald Clahsen. 2013. Brain potentials during language production in children and adults: an ERP study of the English past tense. *Brain & Language*. 127: 345–55.
- Busenitz, Robert and Rene van den Berg. 2012. *A Grammar of Balantak: A Language of Eastern Sulawesi*. Dallas, TX: SIL Publications.
- Bybee, Joan. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. Amsterdam: John Benjamins.
- Carey, Susan. 1978. The child as word learner. In Morris Halle, Joan Bresnan, and George Miller (eds.), *Linguistic Theory and Psychological Reality*, 264–93. Cambridge, MA: MIT Press.
- Ceccagno, Antonella. Undated ms. An Analysis of Chinese Neologisms: Compound Headedness and Classification Issues. University of Bologna.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
1993. A minimalist program for linguistic theory. In Kenneth Hale and Samuel Jay Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 1–52. Cambridge, MA: MIT Press.
1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Clahsen, Harald. 2016. Experimental studies of morphology and morphological processing. In Andrew Hippisley and Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 792–819. Cambridge University Press.
- Clahsen, Harald, Melanie Ring, and Christine Temple. 2004. Lexical and morphological skills in English-speaking children with Williams Syndrome. In Susanne Bartke and Julia Siegmüller (eds.), *Williams Syndrome Across Languages*, 221–44. Amsterdam: John Benjamins.
- Clark, Eve. 2014. Acquisition of derivational morphology. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 424–39. Oxford University Press.
- Comrie, Bernard. 1981/1989. *Language Universals and Linguistic Typology*. 2nd edition. University of Chicago Press.
- Corbett, Greville. 1991. *Gender*. Cambridge University Press.

- Cowell, Mark. 1964. *A Reference Grammar of Syrian Arabic*. Washington, DC: Georgetown University Institute of Language and Linguistics.
- Davies, Mark. 2008. Corpus of Contemporary American English (COCA). <http://corpus.byu.edu/coca/>.
- Davis, Karen. 2003. *A Grammar of the Hoava Language, Western Solomons*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Davis, Stuart and Natsuko Tsujimura. 2014. Nonconcatenative derivation: other processes. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 191–218. Oxford University Press.
- Dayley, Jon. 1985. *Tzutujil Grammar*. Berkeley: University of California Press.
- De Haas, Wim and Mieke Trommelen. 1993. *Morfologisch Handboek van het Nederlands*. The Hague: SDU Uitgeverij.
- Demuth, Katherine. 2000. Bantu noun class systems: loanword and acquisition evidence of semantic productivity. In Gunter Senft (ed.), *Systems of Nominal Classification*, 270–92. Cambridge University Press.
- Desai, Rutvik, Lisa Conant, Eric Waldron, and Jeffrey Binder. 2006. fMRI of past tense processing: the effects of phonological complexity and task difficulty. *Journal of Cognitive Neuroscience* 18, 2: 278–97.
- Diffloth, Gerard. 1976. Expressives in Semai. In Philip Jenner, Laurence Thompson, and Stanley Starosta (eds.), *Austroasiatic Studies*, vol. 2, 249–64. Honolulu: Pacific and Asian Linguistics Institute.
- DiSciullo, Anna Maria and Edwin Williams. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.
- Dixon, R. M. W. 1972. *The Dyirbal Language of North Queensland*. Cambridge University Press.
1977. *A Grammar of Yidiñ*. Cambridge University Press.
- Dressler, Wolfgang. 1985. *Morphonology: The Dynamics of Derivation*. Ann Arbor: Karoma Press.
- Dressler, Wolfgang and Marianne Kilani-Schoch. 2016. Natural Morphology. In Andrew Hippisley and Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 356–89. Cambridge University Press.
- Edmonson, Barbara. 1995. How to become bewitched, bothered, and bewildered: the Huastec version. *International Journal of American Linguistics* 61, 4: 378–95.
- Everett, Daniel and Barbara Kern. 1997. *Wari: The Pacaas Novos Language of Western Brazil*. London: Routledge.
- Fennell, Barbara. 2001. *A History of English: A Sociolinguistic Approach*. Oxford: Blackwell.
- Foley, William. 1991. *The Yimas Language of New Guinea*. Stanford University Press.
- Fortescue, Michael. 1984. *West Greenlandic*. Dover, NH: Croom Helm.
- Fukushima, Kazuhiko. 2005. Lexical V-V compounds in Japanese: lexicon vs. syntax. *Language* 81, 3: 568–612.
- Gaeta, Livio. 2019. Natural Morphology. In Jenny Audring and Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 244–64. Oxford University Press.
- Gagne, Christina and Thomas Spalding. 2019. Morphological theory and psycholinguistics. In Jenny Audring and Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 541–53. Oxford University Press.
- Garrett, Andrew. 2001. Reduplication and infixation in Yurok: morphology, semantics, and diachrony. *International Journal of American Linguistics* 67, 3: 264–312.
- Gibson, Lorna and Doris Barthelemew. 1979. Pame noun inflection. *International Journal of American Linguistics* 45, 4: 309–22.
- Giegerich, Heinz. 1999. *Lexical Strata in English*. Cambridge University Press.
- Goldberg, Adele. 1995. *Constructions*. University of Chicago Press.
- Greenberg, Joseph. (ed.) 1963. *Universals of Language*. Cambridge, MA: MIT Press.
- Guillaume, Antoine. 2008. *A Grammar of Cavineña*. Berlin: Mouton de Gruyter.
- Haas, Mary. 1940. Ablaut and its function in Muskogee. *Language* 16, 2: 141–50.
- Haenisch, Erich. 1961. *Manschu-Grammatik* (Lehrbücher für das Studium der Orientalischen Sprachen 6). Leipzig: VEB Verlag Enzyklopädie.

- Haiman, John. 1980. *Hua: A Papuan Language of the Eastern Highlands of New Guinea*. Amsterdam: John Benjamins.
- Halle, Morris and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale and Samuel J. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 111–76. Cambridge, MA: MIT Press.
- Hardy, Heather and Timothy Montler. 1988. Alabama radical morphology: h-infixation and disfixation. In William Shipley (ed.), *In Honor of Mary Haas*, 377–410. Berlin: Mouton de Gruyter.
- Harley, Heidi. 2009. Compounding in Distributed Morphology. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 129–44. Oxford University Press.
- Harley, Heidi and Rolf Noyer. 1999. Distributed Morphology. *GLoT International* 4, 4: 3–9.
- Harrison, S.P. 1973. Reduplication in Micronesian languages. *Oceanic Linguistics* 12: 407–54.
- Haselow, Alexander. 2011. *Typological Changes in the Lexicon: Analytic Tendencies in English Noun Formation*. Berlin: Mouton de Gruyter.
- Hay, Jennifer and Ingo Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22, 3: 565–96.
- Heine, Bernd. 2014. Areal tendencies in derivation. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 767–76. Oxford University Press.
- Hewitt, B. G. 1979. *Abkhaz*. Amsterdam: NorthHolland.
- Hitchings, Henry. 2005. *Defining the World*. New York: Farrar Straus and Giroux.
- Hoeksema, Jack. 1988. Head-types in morpho-syntax. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology*, vol. 1: 123–38. Berlin: De Gruyter Mouton.
- Hohenhaus, Peter. 2006. Bouncebackability: a web-as-corpus-based case study of a new formation, its interpretation, generalization/spread and subsequent decline. *SKASE* 3, 2: 17–27.
- Horn, Laurence. 2002. Uncovering the un-word: a study in lexical pragmatics. *Sophia Linguistica* 49: 1–64.
- Huddleston, Rodney and Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press.
- Humboldt, Wilhelm von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Königliche Akademie der Wissenschaften.
- Huot, Hélène. 2005. *La Morphologie*. Paris: Armand Colin.
- Inkelas, Sharon. 2014. Non-concatenative derivation: reduplication. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 170–89. Oxford University Press.
- Inkelas, Sharon and C. Orhan Orgun. 1998. Level (non)ordering in recursive morphology: evidence from Turkish. In Steven Lapointe, Diane Brentari, and Patrick Farrell (eds.), *Morphology and Its Relation to Phonology and Syntax*, 360–410. Stanford: CSLI.
- Jaeger, Jeri, Alan Lockwood, David Kemmerer, Robert Van Valin, Brian Murphy, and Hanif Khalak. 1996. A positron tomographic study of regular and irregular verb morphology in English. *Language* 72, 3: 451–97.
- Jeanne, Laverne. 1982. Some phonological rules of Hopi. *International Journal of American Linguistics* 48, 3: 245–70.
- Johnson, Heidi. 2000. *A Grammar of San Miguel Chimalapa Zoque*. Doctoral dissertation, University of Texas at Austin.
- Keenan, Edward and Maria Polinsky. 1999. Malagasy (Austronesian). In Andrew Spencer and Arnold Zwicky (eds.), *The Handbook of Morphology*, 563–623. Oxford: Blackwell.
- Kimball, Geoffrey. 1991. *Koasati Grammar*. Lincoln: University of Nebraska Press.
- Klamer, M. 1998. *A Grammar of Kambara*. Berlin: Mouton de Gruyter.
- Kornfilt, Jaklin. 1997. *Turkish*. London: Routledge.
- Landau, Sidney. 2001. *Dictionaries: The Art and Craft of Lexicography*. 2nd edition. Cambridge University Press.
- Langdon, Margaret. 1970. *A Grammar of Diegueño*. Berkeley: University of California Press.
- Lapointe, Steven. 1980. *A Theory of Grammatical Agreement*. Doctoral dissertation, University of Massachusetts.

- Lederer, Herbert. 1969. *Reference Grammar of the German Language*. New York: Charles Scribner's Sons.
- Lehnert, Martin. 1971. *Rückläufiges Wörterbuch der englischen Gegenwartssprache*. Leipzig: VEB Verlag Enzyklopädie.
- Lewis, G. L. 1967. *Turkish Grammar*. Oxford: Clarendon Press.
- Li, Charles and Sandra Thompson. 1971. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Li, Ya Fei. 1995. The thematic hierarchy and causativity. *Natural Language and Linguistic Theory* 13: 255–82.
- Lieber, Rochelle 1980. On the organization of the lexicon. Doctoral dissertation, Massachusetts Institute of Technology.
1987. *An Integrated Theory of Autosegmental Processes*. Albany: SUNY Press.
1992. *Deconstructing Morphology*. University of Chicago Press.
2004. *Morphology and Lexical Semantics*. Cambridge University Press.
2016. *English Nouns: The Ecology of Nominalization*. Cambridge University Press.
- Lieber, Rochelle and Sergio Scalise. 2007. The Lexical Integrity Hypothesis in a new theoretical universe. *Lingue e Linguaggio* 5, 1: 7–32.
- Lignos, Constantine and Charles Yang. 2016. Morphology and language acquisition. In Andrew Hippisley and Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 765–91. Cambridge University Press.
- Macauley, Monica. 1996. *A Grammar of Chalcatongo Mixtec*. Berkeley: University of California Press.
- Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word Formation: A Synchronic-Diachronic Approach*. Munich: Beck.
- Marshall, Chloe and Heather van der Lely. 2012. Irregular past tense forms in English: how data from children with specific language impairment contribute to models of morphology. *Morphology* 22: 121–41.
- Martin, Jack. 1988. Subtractive morphology as dissociation. *Proceedings of the West Coast Conference on Formal Linguistics* 7: 229–40.
- Massam, Diane. 2009. Noun incorporation: essentials and extensions. *Language and Linguistics Compass* 3, 4: 1076–96.
- McCarthy, John. 1979. Formal Problems in Semitic Phonology and Morphology. Doctoral dissertation, Massachusetts Institute of Technology.
1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12: 373–418.
1983. Phonological features and morphological structure. *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax*, 153–61. Chicago Linguistic Society.
1984. Prosodic organization in morphology. In Mark Aronoff and Richard Oehrle (eds.), *Language Sound Structure*, 299–317. Cambridge, MA: MIT Press.
- Mchombo, Sam. 1999. Chichewa (Bantu). In Andrew Spencer and Arnold Zwicky (eds.), *The Handbook of Morphology*, 500–20. Oxford: Blackwell.
- McGinnis-Archibald, Martha 2016. Distributed Morphology. In Andrew Hippisley and Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 390–423. Cambridge University Press.
- McLaughlin, Fiona. 2000. Consonant mutation and reduplication in Seereer-Siin. *Phonology* 17: 333–63.
- McWhinney, Brian. 2005. Commentary on Ullman et al. *Brain and Language* 93: 239–42.
- Milin, Petar, Laurie Beth Feldman, Michael Ramscar, Peter Hendrix, and Harald Baayen. 2017. Discrimination in lexical decision. *PLOS ONE*. DOI: 10.1371/journal.pone.0171935.
- Mithun, Marianne. 1999. *The Languages of Native North America*. Cambridge University Press.
- Mosel, Ulrike and Even Hovdhaugen. 1992. *Samoan Reference Grammar*. Oslo: Scandinavian University Press.
- Munte, Thomas, Tessa Say, Harald Clahsen, Kolja Schlitz, and Marta Kutas 1999. Decomposition of morphological complex words in English: evidence from event-related brain potentials. *Cognitive Brain Research* 7: 241–53.
- Myers, Scott and Megan Crowhurst. 2006. Case study: sound patterns in Turkish. Retrieved from www.laits.utexas.edu/phonology/turkish/index.html.
- Neef, Martin. 2009. IE, Germanic: German. In Rochelle Lieber and Pavol Štekauer (eds.), *The*

- Oxford Handbook of Compounding*, 386–99. Oxford University Press.
- Newman, Paul. 2000. *The Hausa Language*. New Haven, CT: Yale University Press.
- Nguyen, Hy-Quang and Eleanor Jorden. 1969. *Vietnamese Familiarization Course*. Washington, DC: Foreign Service Institute.
- Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62, 1: 56–119.
- Nida, Eugene. 1946/1976. *Morphology: The Descriptive Analysis of Words*. 2nd edition. Ann Arbor: University of Michigan Press.
- O'Neil, Wayne. 1980. The evolution of the Germanic inflectional systems: a study in the causes of language change. *Orbis* 27: 248–86.
- Peterson, Tyler. 2021. Mirativity in morphology. In Rochelle Lieber (ed.), *The Oxford Encyclopedia of Morphology*, 230–46. Oxford University Press.
- Pinker, Steven. 1999. *Words and Rules*. New York: Perennial.
- Pirrelli, Vito, Ingo Plag, and Wolfgang Dressler. 2020. Word knowledge in a cross-disciplinary world. In Vito Pirrelli, Ingo Plag, and Wolfgang Dressler (eds.), *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon*, 1–20. Berlin: De Gruyter.
- Plag, Ingo. 1999. *Morphological Productivity: Structural Constraints in English Derivation*. Berlin: Mouton de Gruyter.
- Press, Margaret. 1979. *Chemehuevi: A Grammar and Lexicon* (University of California Publications in Linguistics, vol. 92). Berkeley: University of California Press.
- Pulleyblank, Douglas. 1987. Yoruba. In Bernard Comrie (ed.), *The World's Major Languages*, 971–90. New York: Oxford University Press.
- Redmond, Sean and Mabel Rice. 2001. Detection of irregular verb violations by children with and without SLI. *Journal of Speech, Language and Hearing Research* 44: 655–69.
- Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace, Jovanovich.
- Saussure, Ferdinand de. 1972/1983. *Course in General Linguistics*. Translated by Roy Harris. London: Duckworth.
- Schachter, Paul and Fe Otanes. 1972. *Tagalog Reference Grammar*. Berkeley: University of California Press.
- Selkirk, Elisabeth. 1982. *The Syntax of Words*. Cambridge, MA: MIT Press.
- Shaw, Patricia. 1980. *Dakota Phonology and Morphology*. New York: Garland.
- Siddiqi, Daniel. 2019. Distributed Morphology. In Jenny Audring and Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 143–65. Oxford University Press.
- Smith, Norval. 1985. Spreading, reduplication and the default option in Miwok nonconcatenative morphology. In Harry van der Hulst and Norval Smith (eds.), *Advances in Nonlinear Phonology*, 363–80. Dordrecht: Foris.
- Spencer, Andrew. 1991. *Morphological Theory*. Oxford: Blackwell.
- Sridhar, S. N. 1990. *Kannada*. London: Routledge.
- Stockall, Linnea and Alec Marantz. 2006. A single route, full decomposition model of morphological complexity. *The Mental Lexicon* 1, 1: 85–123.
- Stump, Gregory. 2001. *Inflectional Morphology*. Cambridge University Press.
2016. *Inflectional Paradigms*. Cambridge University Press.
2019. Paradigm Function Morphology. In Jenny Audring and Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 285–304. Oxford University Press.
- Suttles, Wayne. 2004. *Musqueam Reference Grammar*. Vancouver: UBC Press.
- Thompson, Sandra, Joseph Sunh-Yul Park, and Charles Li. 2006. *A Reference Grammar of Wappo*. Berkeley: University of California Press.
- Thornton, Anna. 2011. Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology. In Martin Maiden, John Charles Smith, Maria Goldbach, and

- Marc-Olivier Hinzelin (eds.), *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*, 358–81. Oxford University Press.
- Toman, Jindrich. 1983. *Wortsyntax*. Tübingen: Max Niemeyer Verlag.
- Ullman, Michael, Roumyana Pancheva, Tracy Love, Eiling Yee, David Swinney, and Gregory Hickok. 2005. Neural correlates of lexicon and grammar: evidence from the production, reading and judgment of inflection in aphasia. *Brain and Language* 93: 185–238.
- Valentine, J. Randolph. 2001. *Nishnaabemwin Reference Grammar*. University of Toronto Press.
- Velupillai, Viveka. 2012. *An Introduction to Linguistic Typology*. Amsterdam/Philadelphia: John Benjamins.
- Vitale, Anthony. 1981. *Swahili Syntax*. Dordrecht: Foris.
- Watahomigie, Lucille, Jorigine Bender, and Akira Yamamoto. 1982. *Hualapai Reference Grammar*. Los Angeles: American Indian Studies Center, UCLA.
- Wentworth, Harold. 1941. The allegedly dead suffix *-dom* in modern English. *PMLA* 56, 1: 280–306.
- Whaley, Lindsay. 1997. *Introduction to Typology*. Thousand Oaks, CA: Sage Publications.
- Williams, Edwin. 1981. On the notions ‘lexically related’ and ‘head of a word’. *Linguistic Inquiry* 12: 245–74.
- Williams, Marianne Mithun. 1976. *A Grammar of Tuscarora*. New York: Garland.
- Yu, Alan. 2004. Reduplication in English Homeric infixation. In Keir Moulton and Matthew Wolf (eds.), *Proceedings of NELS 34*, 619–33. Amherst, MA: GLSA.

Index

- Abkhaz, 162
ablative, 111
ablaut, 95, 120, 160, 212
-able, 67, 77, 220, 239
absolute, 111, 169
accusative, 111
acquisition, lexical, 17
acronym, 62
adjunct, 167
affix, 37, 38
affixal polysemy, 45, 224
affixation, 39
-age, 44, 220
agent, 44, 107, 168, 247
agglutinative language, 153
agrammatism, 23
agreement, 109, 129
-al, 43, 49, 85, 197
Alabama, 103
Albanian, 159
Alford, Henry, 32
allomorph, 49, 182, 183, 185
allomorphy, 49
American Structuralism, 138
Amharic, 103
analytic language, 152
-ance, 217, 220
-ant, 225, 246
anti-passive, 169
aphasia, 23
 fluent, 23
 non-fluent, 23
apophony, 95. *See* ablaut
applicative, 170
Arabic, 98, 99, 134, 155
areal tendencies, 159
argument, 166
 referential, 247
argument structure, 225
aspect, 113, 114
 completive, 115
 continuative, 115
 habitual, 115, 116
 imperfective, 115
 inceptive, 115
 iterative, 115
 perfective, 115
 quantificational, 115
 semelfactive, 115
assimilation, 183, 184, 195
-ation, 43, 48, 56, 189, 197, 217, 221, 236, 240
augmentative, 44, 129

Baayen, Harald, 16, 77, 78, 79, 80
backformation, 61, 82
Bailey, Nathaniel, 27

Balantak, 187, 188
Bantu, 159
base, 37, 48
 bound, 37, 49, 50, 75
 free, 38
Basque, 112
Berman, Ruth, 19
binarity, 241
binyan, 99
biuniqueness, 242
blending, 62, 82, 83
blocking, 217, 218, 219, 226
Bloom, Paul, 17, 18, 19
bracketing paradox, 221, 226
British English, 28
British National Corpus, 65, 171
Brown, Roger, 22
Bulgarian, 159
Bybee, Joan, 157, 158

Carey, Susan, 18, 19
case, 7, 111, 121, 166
case alignment
 ergative-absolutive, 111
 nominative-accusative, 111
causative, 169, 176
Cavineña, 94
Cawdrey, Robert, 26
Central Alaskan Yup'ik, 12, 114
Central Pomo, 118
Chechen, 118, 162
Chemehuevi, 96
Chichewa, 179
Child, Julia, 84
Chomsky, Noam, 231
circumfix, 93, 94
Clahsen, Harald, 16
Clark, Eve, 19, 20
classifier, 7, 110
clipping, 63
clitic, 112, 172
 simple, 172
 special, 172, 173
closed class, 121
coinage, 61
comparative, 166, 176, 177, 223
competition, 217
Complexity Based Ordering, 221
compound, 50–8, 85, 128, 158
 attributive, 55
 coordinative, 56, 58
 endocentric, 57
 exocentric, 57
 head, 54, 57
 internal structure, 51
 neo-classical, 53

- compound (cont.)
 phrasal, 174
 root, 54
 stress, 51
 subordinative, 56, 58
 synthetic, 54, 56
 verbal, 159
 computational model, 243
 computer learning model, 243
 conjugation, 124
 consonant mutation, 96, 129
 construction, 235
 Construction Grammar (CxG), 235
 Construction Morphology (CxM), 234
 conversion, 58, 95, 241
 corpus, 32, 64, 78
 Corpus of Contemporary American English, 39, 65, 69, 75, 79, 81, 82, 171, 175, 178, 179, 219, 220, 227
 Corpus of Historical American English (COCA), 65, 81, 82, 84, 86
 Covid-19, 83
 cran morph, 47
 cue, 243, 244
 cumulative exponence, 155
 Cupeño, 99
- d, 172
 dative, 111
 de-, 44, 45, 68, 209, 236
 declension, 124
 default ending, 121
 defectiveness, 125, 126, 239, 240
 dependent, 156
 dependent-marking, 156
 derivation, 37, 106, 128, 166, 239
 Developmental Language Disorder. *See* Specific Language Impairment
 dictionary, 11, 13, 25
 Diegueño, 135
 diminutive, 44, 129
 Distributed Morphology (DM), 231
 ditransitive, 167
 Diyari, 192
 -dom, 45, 79, 80, 81, 82, 85, 128, 198, 219
 Dothraki, 204
 double-marking, 157
 Dutch, 19, 45, 57, 59, 77, 93, 135, 173, 174, 175, 200, 203, 224
 Dyirbal, 77, 89, 90, 112, 156, 157
- ed, 21, 24, 113, 120, 121, 128, 168, 182, 222, 238, 242
 -ee, 43, 44, 45, 220, 227
 -eer, 69, 86
 en-, 221
 -en, 76, 77, 168, 221
 enclitic, 172
 Encyclopedia, 233
 English, 2, 36, 46, 47, 48, 50, 60, 74, 75, 76, 80, 83, 88, 106, 115, 125, 126, 128, 149, 154, 155, 160, 176, 178, 179, 208, 209, 215, 217, 224
 ablaut, 95
 acquisition, 19, 22
 acronyms, 63
 affix ordering, 220
 affixation, 39, 40, 44, 45
 allomorphy, 182, 183, 189, 197–200
 aspect, 116
 backformation, 61
 blending, 62, 83
 causative, 176
 clitics, 172
 comparative and superlative, 176
 compounding, 50, 51, 53, 54, 56, 57, 85, 159
 conversion, 58, 69
 corpora, 64
 dependent-marking, 156
 dictionaries, 25–32
 expletive infixation, 92
 formatives, 47
 hypocoristics, 101, 193
 inflection, 7, 106, 119–23
 minor word formation processes, 60
 mood, 118
 number, 106
 passive, 117, 167
 past tense, 21, 22, 113, 184
 phrasal compounds, 174
 phrasal verbs, 173
 plural, 203
 splinters, 48, 69, 86
 umlaut, 96
 epenthesis, 185
 -er, 3, 19, 44, 45, 46, 56, 61, 176, 198, 209, 223, 225, 227, 236, 245, 246, 247
 ergative, 111, 169
 -ery, 44
 -esque, 65, 75, 78, 79, 82, 85, 219
 -ess, 76
 -est, 176
 etymology, 28
 evaluative affix, 44
 event-related potential (ERP), 16, 24
 evidentiality, 118
 experiencer, 225
 extender, 48, 54
 eye tracking, 16
- fast mapping, 18
 figure and ground, 241
 formative, 47
 French, 7, 19, 20, 33, 47, 53, 56, 57, 59, 77, 108, 109, 126, 173, 176, 197, 200, 209
 frequentative, 88, 146
 -ful, 44, 198
 Fula, 44, 129, 130
 functional magnetic resonance imaging (fMRI), 16
 functional shift. *See* conversion
 fusional language, 153
 future, 113
- Gavagai problem, 17
 gender, 108, 110, 121
 natural, 109
 generative grammar, 138, 231
 genitive, 111
 Georgian, 111, 112
 German, 19, 51, 59, 95, 107, 108, 109, 123, 173, 175, 208
 Germanic, 77, 80, 160

- Global Web-Based English (GloWbE), 64
glossing, 112
Greek, 7, 53
Greenberg, Joseph, 157, 158
Gtaʔ, 102
- habilitative, 99
hapax legomenon, 79, 81
Hausa, 96, 109, 133
head, 156, 208
head-marking, 156
Hebrew, 19, 99, 100, 155
Hitchings, Henry, 27
Hoava, 92
-hood, 37, 45, 85, 198, 219, 236
Hopi, 205
Horn, Laurence, 68
Hua, 92
Hualapai, 69, 104, 162
Huastec, 104
Humboldt, Wilhelm von, 152
Hungarian, 156
hyper-, 215
hypo-, 215
hypocoristic, 101, 192
- ian*, 197
-ic, 49, 77, 85, 197, 199
-ical, 37, 49
iconicity, 241
-ie, 44, 101
-ify, 2, 40, 41, 43, 128, 220, 221
in-, 44, 182, 183
index of exponence, 155
index of fusion, 155
index of synthesis, 155
indexicality, 241
Indo-European, 80, 109, 158, 159, 160
infixation, 91–3
inflection, 6, 38, 88, 99, 106, 128, 166, 220
 aspect, 113
 case, 111
 contextual, 127, 158
 English, 119
 evidentiality, 118
 exponence, 155
 implicational universals, 158
 inherent, 127, 158
 inherent *versus* contextual, 127
 mirativity, 118
 mood and modality, 117
 number, 106
 paradigms, 123
 person, 107
 productivity, 126
 syncretism, 125
 tense, 113
 voice, 116
inflectional class, 124
-ing, 14, 56, 120, 126
initialism, 62
instrument, 225
intensive, 146
interfix. *See* linking element
- internal stem change, 88, 95, 120
intransitive, 167
-ish, 38, 85, 198, 199, 219
isolating language, 152
-ist, 246
Italian, 56, 159
Item and Arrangement (IA), 211, 213, 226, 232
Item and Process (IP), 212
-ity, 43, 64, 72, 73, 74, 75, 76, 84, 197, 217, 218, 222, 241
-ive, 197
-ize, 2, 3, 40, 41, 48, 67, 76, 209, 210, 211, 212, 220, 236
- Japanese, 159
Johnson, Samuel, 27
- Kalallisit. *See* West Greenlandic
Kambera, 94
Kannada, 68, 102, 224
Karak, 91
Kersey, John, 26
Koasati, 100, 103
Koyukon, 115
- Landau, Sidney, 26
language acquisition, 22
Late Insertion, 233
Latin, 7, 26, 38, 47, 53, 77, 107, 111, 113, 117, 124, 125, 127, 134, 140, 147–8, 153, 155, 197, 200, 212, 213
learning algorithm, 244
Leipzig glossing rules, 112
-less, 43, 44, 198
-let, 44, 83, 85
Leti, 101, 204
Level Ordering Hypothesis. *See* Stratal Ordering Hypothesis
lexeme, 4, 5, 36, 54
lexeme formation, 6
Lexical Contrast Principle, 18
lexical decision task, 16
Lexical Integrity Hypothesis, 214, 215
Lexical Semantic Framework (LSF), 245
lexical semantics, 247
lexical strata, 197, 199
lexicalization, 31
lexicographer, 13, 26
lexome, 244
Lignos, Constantine, 22
-like, 219
linking element, 54
-ll, 172
locative affix, 44
logographic writing, 33
- magnetoencephalography (MEG), 16, 24
Malagasy, 179
Manchu, 88, 95
Mandarin Chinese, 7, 33, 107, 110, 140, 142–4, 153, 155, 159
Manipuri, 115
Maori, 68, 168
Mapudungun, 171
McWhinney, Brian, 23
mega-, 44, 65, 86
-ment, 43, 56, 217, 221
mental lexicon, 5, 11, 12, 15, 16–24, 189, 208, 219, 230, 234, 235, 237, 239, 247

- mental representation, 234
 Merge, 232
 Middle English, 160
mini-, 82
 Minimalist theory of syntax, 231
 mirativity, 118
 Mithun, Marianne, 12
 Mixtec, 163
 modality, 117. *See* mood
 Mohawk, 107, 108, 171, 216
 Mokilese, 97, 192
mono-, 49
 mood, 117
 declarative, 117
 imperative, 117
 interrogative, 117
 subjunctive, 118
 morpheme, 3, 12, 36, 46, 47
 bound, 37
 free, 37
 mountweazel, 14, 32
 Move, 232
multi-, 44
 multiple exponence, 212
 Murray, James, 28, 29
 Muskogee, 95
 Musqueam, 133, 161, 179
 Mutual Exclusivity Principle, 19

 Na'vi, 136
 Nahuatl, 136, 162
 Naïve Discriminative Learning (NDL), 243
nano-, 82
 Natural Morphology (NM), 240
 naturalness, 240, 241
 language specific, 242
 typological, 242
 negative affix, 44
-ness, 14, 31, 39, 41, 43, 72, 73, 74, 75, 76, 198, 218, 220, 221, 241
 neurolinguistics, 16
 News on the Web, 64, 83
 Nishnaabemwin, 109, 140, 148–51, 154
 Nivkh, 96
 nominative, 111
non-, 44, 48
 nonce, 13, 30
 noun class, 109, 110, 129
 noun incorporation, 171, 216
 number, 6, 106, 122, 166
 dual, 107

 Ojibwa, 109
 Old English, 7, 122, 123, 125, 154, 160, 161, 198
 Old Norse, 123
-ory, 197
-ous, 217
out-, 44, 178
 outcome, 244
over-, 44, 171
 overabundance, 125, 126, 239, 240
 Oxford English Dictionary, 13, 28, 34, 80, 84, 85

 palatalization, 190
 Pame, 205

 paradigm, 123, 234, 238
 derivational, 240
 Paradigm Function, 238, 239
 Paradigm Function Morphology (PFM), 238
 parasyntesis, 93
 participant affix, 44
 passive, 167
 patient, 44, 168
 perfect, 120
 periphrasis, 176
 person, 7, 106, 107, 122, 166
 exclusive, 108
 inclusive, 108
 phrasal compound, 215
 phrasal verb, 173, 174
 Pinker, Steven, 22
 polysynthetic language, 154, 171
 portmanteau word. *See* blending
 positron emission tomography (PET), 16, 24
post-, 44, 215
pre-, 3, 44, 215
 prefix, 37, 38, 48
 prepositional affix, 44
 priming, 16
 Principles and Parameters, 231
 privative affix, 44
 proclitic, 172
 productivity, 73, 126
 compositionality, 75
 creativity, 82
 frequency of base type, 75
 lexicalization, 75, 79
 measurement of, 78
 transparency, 74, 78, 79, 85
 usefulness, 76
 progressive, 120
 psycholinguistics, 16

 quantitative affix, 44
 Quechua, 162
 Quine, W.O., 17

re-, 3, 12, 42, 43, 44, 68
 reaction time, 16, 245
 realis/irrealis, 117
 Realization Rule, 238
 realizational model, 212
 reduplication, 88, 96–8, 190
 echo, 97
 full, 97
 partial, 97, 190
 relational affix, 44
 response latency, 16
 Righthand Head Rule, 208
 rival affixes, 217
 root, 38
 root and pattern morphology. *See* templatic morphology
 Romanian, 159
 Rule of Exponence, 238
 Rule of Referral, 238
 Russian, 7, 134

-s, 7, 106, 119, 121, 128, 242
 Samoan, 88, 97, 140, 144–6

- Sapir, Edward, 139
 schema, 236
 second order, 237
 Seereer-Siin, 96
 Semai, 103
 semantic body, 246
 semantic features, 246
 Seneca, 115
 separable prefix verb, 174
 Sesotho, 129
 -ship, 85, 219
 Sierra Miwok, 99
 Simpson, Homer, 103
 simulfix. *See* internal stem change
 skeleton, 246
 slang, 34
 Sm'algyax, 112
 -some, 83
 Sorbian, 125, 239
 sound symbolism, 48
 Spanish, 7, 56
 Specific Language Impairment (SLI), 23
 speech act, 117
 Spell Out, 233
 splinter, 48, 86
 stem, 38, 39
 storage versus rules, 245
 Stratal Ordering Hypothesis, 220, 227
 Stump, Gregory, 239
 subtractive morphology, 100–1, 241
 suffix, 2, 37, 38, 48
 superlative, 166, 176, 177
 suppletion, 125
 Swahili, 7, 134, 161, 170, 176, 178, 209
 Swedish, 19
 syllable, 190
 coda, 190
 heavy, 192
 light, 192
 nucleus, 190
 onset, 190
 rhyme, 190
 syncretism, 125, 126, 239
 synonymy, 218
 System Adequacy, 242

 Tagalog, 88, 91, 93, 133, 193, 202
 template, 99
 templatic morphology, 88, 98–100, 155, 226
 tense, 6, 106, 113
 on nouns, 114
 periphrastic marking, 113
 Teton Dakota, 97
 -th, 38, 72, 73, 82, 221
 theme, 168, 226
 theme vowel, 124
 token, 78
 Tonkawa, 117
 transfix. *See* templatic morphology
 transitive, 167
 transparency, 74, 75, 241
 transposition, 72

 transpositional affixes, 43
 trilateral root, 99
 Turkish, 113, 133, 140–2, 153, 154, 155, 182, 185, 186, 213, 215, 224
 Tuscarora, 44
 type, 78
 typology, 138, 152, 238, 243
 Tzutujil, 94, 157, 162, 204

 Ulwa, 92
 umlaut, 96, 120
 un-, 19, 37, 38, 39, 40, 41, 44, 68, 76, 222
 underlying representation, 183
 Unitary Base Hypothesis, 209
 universals, 138, 139
 implicational, 158
 -ure, 217

 valency, 166, 167
 Vietnamese, 53, 57, 152, 153, 209
 Vocabulary Item, 233
 voice, 116
 active, 116
 passive, 117
 voicing assimilation, 184
 vowel harmony, 140, 186, 188

 Wappo, 201
 Wari', 155
 Washo, 114
 Webster, Noah, 27, 28
 West Greenlandic, 115, 168
 Whole Object Principle, 19
 Wiktionary, 32
 Williams Syndrome, 23
 Wintu, 118
 Worcester, Joseph, 28
 word, 3, 12, 25
 complex, 3
 simplex, 3
 Word and Paradigm, 212, 213
 word formation, 2. *See also* lexeme formation
 word formation rule, 17, 40, 41
 word forms, 6
 word token, 4, 5
 word tree, 42, 52
 word type, 4, 5
 words *versus* rules, 20, 22, 23
 World Atlas of Linguistic Structures, 152, 155

 -y, 44, 101
 Yang, Charles, 22
 Yay, 161
 Yiddish, 97
 Yidip, 169
 Yimas, 130, 131, 132
 Yoruba, 45, 224
 Yuchi, 110
 Yurok, 102

 zero affixation. *See* conversion
 Zoque, 91, 184, 194

